The confounding effects of high genetic diversity on the determination and interpretation of differential gene expression analysis in the parasitic nematode Haemonchus contortus

Elsevier Editorial System(tm) for
International Journal for Parasitology
Manuscript Draft

Manuscript Number: IJPara19_132R1

Title: The confounding effects of high genetic diversity on the determination and interpretation of differential gene expression analysis in the parasitic nematode Haemonchus contortus

Article Type: Full Length Article

Keywords: Haemonchus contortus; Transcriptomics; RNAseq; Differential Expression; Ivermectin; Anthelmintic Resistance

Corresponding Author: Professor John Gilleard, BVSc PhD DipEVPC MRCVS

Corresponding Author's Institution: University of Calgary, Fac. Vet. Med.

First Author: Andrew M Rezansoff

Order of Authors: Andrew M Rezansoff; Roz Laing; Axel Martinelli; Susan Stasiuk, PhD; Elizabeth Redman; Dave Bartley; Nancy Holroyd; Eileen Devaney; Stephen Doyle; Neil D Sargison; James Cotton; John S Gilleard

Manuscript Region of Origin: CANADA

Abstract: Differential expression analysis between parasitic nematode strains is commonly used to implicate candidate genes in anthelmintic resistance or other biological functions. We have tested the hypothesis that the high genetic diversity of an organism like Haemonchus contortus could complicate such analyses. First, we investigated the extent to which sequence polymorphism affects the reliability of differential expression analysis between the genetically divergent H. contortus strains MHco3(ISE), MHco4(WRS) and MHco10(CAVR). Using triplicates of 20 adult female worms from each population isolated under parallel experimental conditions, we found that high rates of sequence polymorphism in RNAseq reads were associated with lower efficiency read mapping to gene models under default TopHat2 parameters, leading to biased estimates of inter-strain differential expression. We then showed it is possible to largely compensate for this bias by optimizing the read mapping SNP allowance and filtering out genes with particularly high SNP rates. Once the sequence polymorphism biases were removed, we then assessed the genuine transcriptional diversity between the strains, finding ≥ 824 differentially expressed genes across all three pairwise strain comparisons. This high level of inter-strain transcriptional diversity not only suggests substantive inter-strain phenotypic variation but also highlights the difficulty of reliably associating differential expression of specific genes with phenotypic differences. To provide a practical example, we analyzed two gene families of potential relevance to ivermectin drug resistance; the ABC transporters and the ligand-gated ion channels (LGICs). Over half of genes identified as differentially expressed using default TopHat2 parameters were shown to be an artifact of sequence polymorphism differences. This work illustrates the need to account for sequence polymorphism in differential expression analysis. It also demonstrates that a large number of genuine transcriptional differences can occur between H. contortus strains and these must be

considered before associating the differential expression of specific genes with phenotypic differences between strains.

***Cover Letter**

UNIVERSITY OF CALGARY

John Gilleard, BVSc, PhD, DipACVM, DipEVPC, MRCVS
Associate Dean, Research
Professor, Molecular Parasitology
Faculty of Veterinary Medicine
TRW 2D10, 3280 Hospital Drive NW
Calgary, Alberta T2N 4Z6

Telephone: +1-403-210-6327
Email: jsgillea@ucalgary.ca

May 17, 2019

Dear Dr Loukas,

Thank you for your consideration of our manuscript IJPara19_132 for publication in the International Journal for Parasitology entitled *"The confounding effects of high genetic diversity on the determination and interpretation of differential gene expression analysis in the parasitic nematode Haemonchus contortus"*. We appreciate the positive evaluation of our manuscript and thank you and the referees for the constructive feedback provided. We submit a revised manuscript, together with a line-by-line response to the reviewers' comments

We hope the changes made to this section are satisfactory.

We thank you for considering this revised manuscript and hope it is now considered suitable for publication.

Yours sincerely,

Prof John Gilleard

***Response to Reviewers**

John Gilleard, BVSc, PhD, DipACVM, DipEVPC, MRCVS
Associate Dean, Research
Professor, Molecular Parasitology
Faculty of Veterinary Medicine
TRW 2D10, 3280 Hospital Drive NW
Calgary, Alberta T2N 4Z6

Telephone: +1-403-210-6327
Email: jsgillea@ucalgary.ca

May 17, 2019

Dear Dr Loukas,

Thank you for your consideration of our manuscript IJPara19_132 for publication in the International Journal for Parasitology entitled *"The confounding effects of high genetic diversity on the determination and interpretation of differential gene expression analysis in the parasitic nematode Haemonchus contortus"*. We appreciate the positive evaluation of our manuscript and thank you and the referees for the constructive feedback provided. We submit a revised manuscript, together with a line-by-line response to the reviewers' comments and an itemized list of modifications to the revised manuscript. The reviewer comments are included in *italics* and our response in non-italicized text. The line numbers referred to in our responses refer to those of the revised manuscript.

Reviewer 1

We thank the reviewer for their very positive view and constructive feedback of the manuscript.

We have addressed the reviewers points as follows:

Reviewer comment 1: Abstract: lines 29-37. It is important to have some statement in here about what lifestage(s), age, sex, host, method of worm collection etc were used to ensure that it's a level playing field at the start - e.g. that you're not comparing MHco3 L3s that have been kept in a fridge for a year with adult female MHco10 straight out of an abomasum.

Response: Line 32 - We have inserted a comment describing the experimental samples to clarify the controlled nature of RNAseq experiment from the abstract onward. We also made some minor grammatical changes to the abstract to maintain its length below 300 words again.

Reviewer comment 2: Figure 2C, the labelling on the x axis could be clearer - i.e. leave more room between the labels

Response: We have condensed Figure 2C X-axis labels to increase the space between them as requested.

Reviewer 2

We have addressed the reviewers points as follows:

Reviewer comment 1: This interesting manuscript makes the case that simply using the default setting on the most commonly used suite of programs for the analysis of RNASeq data can lead to serious errors in the estimation of changes in gene expression between strains and isolates. The authors specifically looked at Haemonchus contortus, though presumably the point is also valid for many other parasites in which drug resistance, or another important biological variable, is a problem.

Response: We thank the reviewer for their positive comments

Reviewer comment 2: With such papers it is essentially impossible to judge the quality of the experimental approach; this is an excellent and reputable group so I would not expect there to be any problems.

Response: We thank the reviewer for their supportive comments

Reviewer comment 3: The authors cite several previous reports in their critique of the existing literature, however at least some of these use qPCR rather than RNASeq data, and so the criticism may not be valid. Would they expect qPCR to be equally prone to these artefactual differences between strains - perhaps not assuming that the primer sequences are conserved. If that were not the case it might be that the amplifications would fail? Some discussion of this point might be valuable.

Response: Line 391 – We understand the reviewer's point here and have added a comment to the Discussion section to make it more clear that we are suggesting high genetic diversity is a problem when conducting RNAseq experiments specifically.

Reviewer comment 4: This may be an odd statement after the previous point, but overall I found the discussion to be overlong and repetitive of points made elsewhere - it could be considerably shortened with no loss of impact.

We would prefer to retain the current Discussion as it draws together the key messages from the Results. The paper is quite technical in nature and a shortened high-level Discussion would make it difficult for readers to follow how the conclusions were supported.

Reviewer comment 5: The number of reads mapping to the H. contortus genome assembly seems a little low (60-70%, depending on strain). Did the authors check for host or bacterial contamination of their sequences?

Response: The mapping success rates of the samples we observed for all three strains were similar to mapping statistics of other *H*. *contortus* samples that we and others have worked with in the past

(~50-75% mapping success rates is typical for *H. contortus*). Further our map rates among the three triplicates of the same strain were very similar for all three strains, which reasonably suggests that there is no single triplicate that has a low mapping success rate because of contamination.

Reviewer comment 6: Some of the values on the axes of Fig 3 are rather small and faint, and may not reproduce well especially if they are reduced in size.

Response: We have increased the font size of the number labels on both axis of Figure 3 as requested.

Reviewer comment 7: Though 'N' is defined in the Methods, this may be easy for the more casual reader to overlook. Given its central importance to the arguments in the manuscript, it may be worth restating what this value represents in the Results or Discussion, perhaps relating it to the observed level of polymorphism between H. contortus strains.

Response: Lines 209, 213, 214 – We have redefined what N2, N5, and N10 represent when they are first mentioned in the Results. N2 and N5 were also redefined on Lines 411 and 412 of the Discussion section.

We hope the changes made to this section are satisfactory.

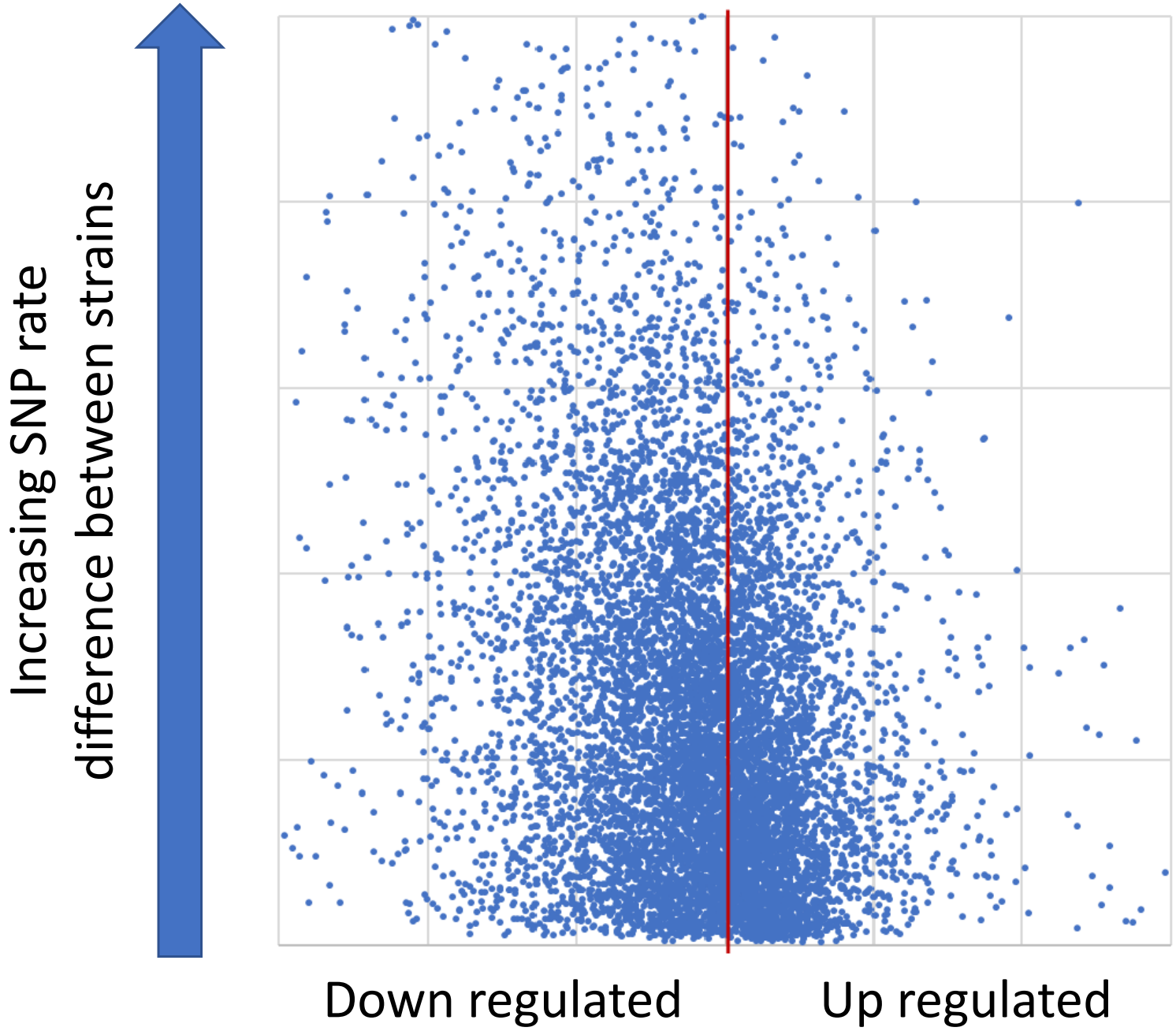
We thank you for considering this revised manuscript and hope it is now considered suitable for publication.

Yours sincerely,

Prof John Gilleard

*Graphical Abstract (for review)

Artifactual bias in differential expression calls due to sequence polymorphism



***Highlights (for review)**

Paper Highlights

- Sequence polymorphism can counfound RNAseq analysis in genetically diverse organisms due to read mapping biases
- Optimizing read mapping allowances and excluding highly polymorphic genes reduces differential gene expression analysis biases
- Genetically divergent strains of *H. contortus* have very high levels of inter-strain transcriptional diversity
- Interpretation of inter-strain differential gene expression needs to consider sequence polymorphism and overall transcriptional diversity

The confounding effects of high genetic diversity on the determination and interpretation of differential gene expression analysis in the parasitic nematode *Haemonchus contortus*

Andrew M. Rezansoff[a], Roz Laing[b], Axel Martinelli[c], Susan Stasiuk[a], Elizabeth Redman[a], Dave Bartley[d], Nancy Holroyd[c], Eileen Devaney[b], Neil D. Sargison[e], Stephen Doyle[c], James Cotton[c], John S. Gilleard[a*]

[a]*Department of Comparative Biology and Experimental Medicine, Faculty of Veterinary Medicine, University of Calgary, Alberta, Canada*

[b]*Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Scotland, United Kingdom*

[c]*Wellcome Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, United Kingdom*

[d]*Moredun Research Institute, Pentlands Science Park, Bush Loan, Penicuik EH26 0PZ, United Kingdom*

[e]*University of Edinburgh, Royal (Dick) School of Veterinary Studies, Easter Bush Veterinary Centre, Roslin, Midlothian, EH25 9RG, United Kingdom*.

* Corresponding author: Professor John Gilleard, Department of Comparative Biology and Experimental Medicine, Faculty of Veterinary Medicine,
3330, Hospital Drive, University of Calgary, Calgary, Alberta, T2N 4N1 Canada
Tel: (403) 210 6327, Fax: (403) 210 7882, email: jsgillea@ucalgary.ca

Abstract

Differential expression analysis between parasitic nematode strains is commonly used to implicate candidate genes in anthelmintic resistance or other biological functions. We have tested the hypothesis that the high genetic diversity of an organism like *Haemonchus contortus* could complicate such analyses. First, we investigated the extent to which sequence polymorphism affects the reliability of differential expression analysis between the genetically divergent *H. contortus* strains MHco3(ISE), MHco4(WRS) and MHco10(CAVR). Using triplicates of 20 adult female worms from each population isolated under parallel experimental conditions, we found that high rates of sequence polymorphism in RNAseq reads were associated with lower efficiency read mapping to gene models under default *TopHat2* parameters, leading to biased estimates of inter-strain differential expression. We then showed it is possible to largely compensate for this bias by optimizing the read mapping SNP allowance and filtering out genes with particularly high SNP rates. Once the sequence polymorphism biases were removed, we then assessed the genuine transcriptional diversity between the strains, finding ≥ 824 differentially expressed genes across all three pairwise strain comparisons. This high level of inter-strain transcriptional diversity not only suggests substantive inter-strain phenotypic variation but also highlights the difficulty of reliably associating differential expression of specific genes with phenotypic differences. To provide a practical example, we analyzed two gene families of potential relevance to ivermectin drug resistance; the ABC transporters and the ligand-gated ion channels (LGICs). Over half of genes identified as differentially expressed using default *TopHat2* parameters were shown to be an artifact of sequence polymorphism differences. This work illustrates the need to account for sequence polymorphism in differential expression analysis. It also demonstrates that a large number of genuine

transcriptional differences can occur between *H. contortus* strains and these must be considered before associating the differential expression of specific genes with phenotypic differences between strains.

Keywords: *Haemonchus contortus*; Transcriptomics; RNAseq; Differential Expression; Ivermectin; Anthelmintic Resistance

1. Introduction

RNAseq has become the standard approach for the genome-wide analysis and quantification of gene expression across the life sciences (Conesa et al., 2016; Wang et al., 2009). Established sequence aligners used in RNAseq analysis pipelines, such as *TopHat2* and its faster successor *HISAT2* were developed, and their default mapping parameters set, primarily for use on vertebrate species such as humans, mouse, and zebrafish, which have relatively low levels of both intra- and inter-population genetic diversity (Baruzzo et al., 2017; Guryev et al., 2006; Lindblad-Toh et al., 2000; Wang, 1998). Further, until relatively recently, applications of RNAseq to non-vertebrate species were largely confined to laboratory strains of model organisms such as *Drosophila melanogaster* and *Caenorhabditis elegans*, which also have relatively low levels of genetic diversity (Andersen et al., 2012; Cingolani et al., 2012). Consequently, most publications make little or no acknowledgement of the potentially confounding effects of sequence polymorphism on the mapping efficiency of RNAseq reads and the calling of differentially expressed genes (Baruzzo et al., 2017). RNAseq analysis pipelines are generally applied to non-model organisms simply using established default parameters, with no consideration given the level and distribution of sequence polymorphism within, and between the strains or populations being compared (Antony et al., 2016; Croken et al., 2014; Edwards et al.,

2013; Fiebig et al., 2015; Papenfort et al., 2015). However, many taxa show high levels and complex patterns of intra-species genetic diversity (Blumenthal and Davis, 2004; Dey et al., 2013; Redman et al., 2015; Romiguier et al., 2014). This is a concern since standard RNAseq alignment benchmarking studies have shown that the performance of different sequence aligners varies with the genome complexity and levels of sequence polymorphism when using simulated sequence data (Baruzzo et al., 2017). However, no published experimental studies directly examine the effects of sequence polymorphism on differential expression analyses using commonly applied RNAseq analysis pipelines.

A good example of the application of RNAseq analysis to non-model organisms is for the investigation of differential expression of candidate genes potentially involved in anthelmintic drug resistance in parasitic nematodes (Dicker et al., 2011; El-Abdellati et al., 2011; Urdaneta-Marquez et al., 2014; Williamson et al., 2011; Xu et al., 1998). *Haemonchus contortus* is arguably the most established parasitic nematode model used for such studies (Gilleard, 2013). It has a good quality reference genome and has extremely high levels of sequence polymorphism (upwards of 5% SNP rates), both within and between strains or geographical isolates (Gilleard and Redman, 2016; Laing et al., 2013). Consequently, it is an excellent system in which to study the potentially confounding effects of sequence polymorphism on differential expression analysis. In this paper, we use three well characterized laboratory passaged strains of *H. contortus* to examine how differences in coding sequence (CDS) polymorphism rates, with respect to the MHco3(ISE) genome reference strain, affect read mapping and bias differential expression analysis. We show how these confounding effects can be reduced and demonstrate that, even when the effects of sequence polymorphism are minimized, there are still a large number of differentially expressed genes between these three strains. These results have important implications for the application of RNAseq analysis to many non-model organism species with high levels of genetic diversity.

2. Materials & Methods

2.1 H. contortus strains, sample preparation, and sequencing.

The MHco3(ISE), MHco4(WRS) and MHco10(CAVR) *H. contortus* strains have been previously characterised and are described in detail elsewhere (Laing et al., 2013; Redman et al., 2012, 2008). The MHco3(ISE) is susceptible to all main classes of anthelmintic and has been used as the reference genome strain (Laing et al., 2013). The MHco4(WRS) strain is derived from the White River Strain (WRS) that was isolated as an ivermectin resistant field isolate from South Africa (Van Wyk and Malan, 1988). The MHco10(CAVR) strain is derived from the Chiswick Avermectin Resistant Strain (CAVR) which was originally isolated as an ivermectin resistant strain as a laboratory contaminant of a field isolate from Australia (Le Jambre et al., 1995).

Three sets of 20 adult female worms were recovered on necropsy at 28 days post experimental infection from the abomasa of three different individual sheep for each *H. contortus* strain; MHco3(ISE), MHco4(WRS), and MHco10(CAVR). Each set of 20 adult females served as one of three biological replicates for RNAseq analysis for each strain. Adult worms recovered from the abomasum were rinsed and sexed in physiological saline at 37^{o}C and then immediately snap frozen before total RNA was isolated from each pool of 20 worms using a standard Trizol protocol as described in Laing et al., (2011). RNA samples were assessed on a Bioanalyser 2100 (Agilent) and Illumina transcriptome libraries were prepared as previously described (Laing et al., 2011). Sequencing of transcriptome libraries was performed on an Illumina HiSeq platform to generate 100 bp paired-end reads.

2.2 Sequence quality control and read mapping.

Raw 100 bp reads were inspected using *FastQC* (Andrews, 2010) for overall dataset integrity and all reads were trimmed at the 5' end by ten bases. Fifteen bases were also trimmed from the 3' ends of all reads to remove low quality sequence characteristic of 3' tail ends. The post-trimmed 75 base-pair reads were used for mapping to the *H. contortus* MHco3(ISE) reference genome assembly (Laing et al., 2013) with *TopHat2* (Dobin and Gingeras, 2013). The assembly used is an improved version (N50 of 5.24 MB) of the original published *H. contortus* genome assembly (GenBank ID PRJEB506 - N50 of 83.29 kb (Laing et al., 2013)) and contains an expanded set of annotated gene models (https://data.mendeley.com/drafts/4z6xv5j5zf). Numerical identifiers of these additional gene models begin with HCOI_0500, and have not yet been submitted to online genomic resources (e.g. *Uniprot.org*).

TopHat2 was executed using the following parameter settings: *TopHat2 -N (#) --read-gap-length (%) --read-edit-dist (# + %) -I 40000 -r 200 -a 6 -g 1 --no-discordant --no-mixed --min-intron 10 --microexon-search --mate-std-dev 50 --library-type fr-unstranded ./reference.fasta trimmed_forward_reads.fastq trimmed_reverse_reads.fastq*. Only *-N* (specifying the number of SNPs per mapped read allowed by *TopHat2*), *--read-gap-length* (the allowed base count of any indels), and *--read-edit-dist* (the allowed combined base count of both *-N* and *--read-gap-length*) were adjusted throughout the experiment. Reads of all triplicates of all three populations were initially mapped with *TopHat2* using a scale of SNP (polymorphism) allowances from 2 to 10 SNPs (*-N*) per read with indel allowance (*--read-gap-length*) held constant at 3 bases.

Three different allowances for polymorphism were then subsequently chosen for further analysis: low, the *TopHat2* default allowances (denoted N2 – allowing 2 SNPs or 2 indels per read), moderate (denoted N5 - allowing 5 SNPs and 3 indels per read), and high (denoted N10 - allowing 10 SNPs and 6 indels per read) allowances for polymorphism respectively. Varying the indel allowances had very little effect on the percentage of reads mapping to the reference genome (data not shown). *Samtools' flagstat* tool (Li et al., 2009) was used to determine the proportion of reads mapped at each allowance for each strain.

2.3 RNAseq processing and analysis.

Reads mapped to each gene model were sorted with *samtools sort*, and counted for each of the three bioreplicates for each strain at the three different SNP allowances – N2, N5, N10 – using the following command in *HTseq-count*: *htseq-count -i parent -q -s no -f bam -t cds ./sorted_accepted_hits.bam ./genome_annotation_file.gff3* (Anders et al., 2014). Raw mapped read counts for each gene model of each bioreplicate of each strain were compiled and used as input for *DESeq2*.

DESeq2 (Love et al., 2014) was run in *Rstudio* (2015) to identify differential expression between the three strains, at different polymorphism allowances, based on gene model read counts. *DESeq2's plotPCA* tool was used to plot segregation of triplicates based on gene expression of the top 15,000 expressed low-polymorphic genes at the moderate N5 allowance. *DESeq2* result tables were exported and manipulated in *MS Excel*. Genes were only called as differentially expressed in this analysis if they 1) showed a greater than 2 fold-change difference in expression between the strains compared, and 2) yielded adjusted p-values of less than 0.05.

161

162 *2.4 Categorizing gene models on the basis of SNP rates and SNP rate differences between strains*

163

164 SNPs within coding regions (CDS) were called using *samtools mpileup* on whole genome
165 sequence (WGS) datasets created for each of the strains against the MHco3(ISE) genome assembly
166 (Doyle et al., 2019). SNPs present at > 40% frequency were totaled per gene model for each of the
167 strains. The SNP rate was calculated for each gene in each strain by dividing the total number of SNPs
168 in the gene by the respective gene model CDS length. The genes were then categorized in two different
169 ways for subsequent investigation of the effect of sequence polymorphism on read mapping and
170 RNAseq analysis. First, they were categorized based on their SNP rates in each strain: categories 0%,
171 0-0.5%, 0.5-1%, 1-2%, 2-5%, and > 5%. Second, they were categorized based on the difference in SNP
172 rates for each of the three pairwise strain comparisons (i.e. the SNP rate observed in one strain
173 subtracted by the SNP rate observed in the other) categories >5-15%, >2-5%, >0-2%, 0%. Genes with a
174 >15% difference and were not categorized as they were likely to be due to annotation errors and/or
175 overly short CDS lengths.

176

177 *2.5 Assessment of genuine transcriptomic variation between the strains*.

178

179 Differential expression statistics were called with *DESeq2* for each of the three pairwise strain
180 comparisons at each of the three map allowances. In each pairwise strain comparison at the N5
181 allowance, genes showing low SNP rate differences (less than 2%) were denoted as low-polymorphic
182 genes (LPGs). The number of low-polymorphic genes up- and down-regulated in each strain
183 comparison at the N5 allowance, and shared up- or down-regulated in two strains vs. the third strain,

were totaled at both a log2 1X and log2 2X fold-change expression threshold. Candidate anthelmintic resistance gene families, as defined by the published *H. contortus* genome annotation (Laing et al., 2013), were specifically highlighted in that their differential expression was compared at the N2 allowance, the N5 allowances, and the N5 allowance with high-polymorphic genes removed.

Gene ontological classifications were obtained from *UniProt.org* (The UniProt Consortium, 2015) for *H. contortus* gene models of the originally published annotation (Laing et al., 2013). Low polymorphic genes with ontological classifications were used as the reference gene set against which enrichment was assessed. Functional enrichment was called in genes > log2 1X fold-change differentially expressed in each pairwise, and each shared strain comparison. *FunRich* (Pathan et al., 2015) was used to call enriched gene ontological classes using a statistical significance threshold of Benjamini-Hochberg corrected FDR adjusted p-values < 0.05.

3. Results

3.1 Coding sequence polymorphism affects RNAseq read mapping against the MHco3(ISE) reference assembly for the three different H. contortus strains.

The total combined read counts of the triplicate RNAseq datasets were similar among the three strains at 36,175,121, 36,025,170, and 37,584,775 reads for MHco3(ISE), MHco4(WRS), and MHco10(CAVR) respectively. We determined the total number of CDS SNPs present at > 40% frequency, relative to the MHco3(ISE) reference genome assembly, using whole genome sequence datasets independently created for each strain. A total of 701,715, 1,121,242 and 1,143,102 CDS SNPs,

representing rates of 2.97%, 4.74% and 4.84% of the 23.63 Mb *H. contortus* reference CDS annotation, were present for MHco3(ISE), MHco4(WRS), and MHco10(CAVR) respectively.

The percentage of RNAseq reads that mapped to the MHco3(ISE) reference genome assembly, using the default SNP allowance (N2 – allowing 2 SNPs or 2 indels per read) in *TopHat2*, was 60.7%, 44.8% and 47.1% for the MHco3(ISE), MHco4(WRS) and MHco10(CAVR) strains respectively (Fig. 1). Increasing the *TopHat2* SNP allowance parameter changed the percentage of RNAseq reads that mapped (Fig. 1). For the MHco3(ISE) strain, the percentage of RNAseq reads mapping to the reference genome increased as the polymorphism allowance was increased from N2 to N5 (allowing 5 SNPs and 3 indels per read) and then decreased as the allowance was further increased to N10 (allowing 10 SNPs and 6 indels per read) (Fig. 1). This pattern was very similar for the MHco4(WRS) and MHco10(CAVR) strains but the maximum percentage of reads mapping occurred at the N6 allowance, albeit at rates only 0.1% greater than at N5 (Fig. 1). The percentage of RNAseq reads that mapped to the reference MHco3(ISE) genome assembly was greater for the MHco3(ISE) strain than for the other two strains at all polymorphism allowances, although the magnitude of this difference decreased from the N2 to N10 allowance (Fig. 1).

A more detailed analysis was undertaken for the N2, N5 and N10 polymorphism allowances at the level of gene models. Increasing the polymorphism allowance from N2 to N5 resulted in 12,778, 11,101, and 11,324 gene models having a >1% increase in the number of mapped RNAseq reads for MHco3(ISE), MHco4(WRS), and MHco10(CAVR) respectively (Fig. 2A, panel i). In contrast, 591, 1,316, and 1,563 genes showed a >1% decrease in RNAseq reads mapped (Fig. 2A, panel i). Further increasing the mapping allowance from N5 to N10 had the opposite effect, with a greater number of gene models having a decreased rather than an increased number of RNAseq reads mapped: A change in the polymorphism allowance from N5 to N10 resulted in 12,529, 8,139, and 8,470 gene models

having a >1% decreased number of RNAseq reads mapped, compared with 1,092, 4,682 and 4,953 genes having an increased number of RNAseq reads mapped for MHco3(ISE), MHco4(WRS), and MHco10(CAVR) strains respectively (Fig. 2A, panel ii).

3.2 The SNP allowance has a greater effect on RNAseq read mapping for gene models with higher levels of sequence polymorphism.

There were large differences in the SNP rates of different gene models, relative to the MHco3(ISE) reference genome, ranging from those with SNP rates of 0% to those above 5%. The 25,111 gene models were binned into several different SNP rate categories to investigate how the mapping of RNAseq reads to the reference MHco3(ISE) genome assembly was affected by the coding region SNP rate (Fig. 2B). The MHco4(WRS) and MHco10(CAVR) strains had a significantly greater proportion of gene models with SNP rates greater than 0.5% [18,910 (75.3%) and 18,886 (75.2%) respectively] compared with the MHco3(ISE) strain [11,303 (45.0%)] (Z-stat = 69.3 ($p < 0.000$) and 69.1 ($p < 0.000$) respectively) (Fig. 2B).

The effect of changing the polymorphism allowance from N2 to N5 on RNAseq read mapping for each of the different SNP rate categories of gene models was examined for each strain (Fig. 2C, panel i; Supplementary Table S1). The ratio of RNAseq reads mapping to gene models at the N5 compared to the N2 allowance was > 1 for all SNP rate categories above 0% for all three strains (Fig. 2C, panel i). Furthermore, this ratio increased as the SNP rate increased. In contrast, the ratio of RNAseq reads mapping to gene models at the N10 allowance compared to the N5 allowance was < 1 except for gene models with a polymorphism frequency of > 5% for strains MHco4(WRS) and MHco10(WRS) (Fig. 2C, panel ii).

252

253 *3.3 High levels of sequence polymorphism artificially inflate between-strain RNAseq differential*
254 *expression results*.

255

256 We next investigated the influence of CDS sequence polymorphism on the RNAseq differential
257 expression reported by *DESeq2* between pairwise strain comparisons. We hypothesized that gene
258 models with large differences in SNP rates (SNPs/bp) between two strains are more likely to be
259 reported as differentially expressed between those strains than gene models with smaller SNP rate
260 differences. To test this hypothesis, for each gene model we first determined the difference in SNP
261 rates (SNPs/bp) between each pairwise comparison of the three strains. We then plotted the difference
262 in the SNP rate between the two strains against the log2-fold difference in expression called by *DESeq2*
263 for each gene model (Fig. 3). Using the MHco4(WRS) and MHc03(ISE) pairwise comparison as an
264 example, for those gene models with a higher SNP rate in MHco4(WRS) than in MHco3(ISE), a
265 greater number was reported by *DESeq2* as down-regulated in MHco4(WRS) relative to MHco3(ISE)
266 than as up-regulated (Fig. 3A). This bias towards down-regulation increased as the SNP rate difference
267 of gene models between the two strains increased (Fig. 3A). For gene models with a lower SNP rate in
268 MHco4(WRS) than in MHco3(ISE), the opposite trend was apparent (Fig. 3B). Similar patterns were
269 observed in both the MHco3(ISE) vs. MHco10(CAVR) and MHco4(WRS) vs. MHco10(CAVR)
270 pairwise comparisons (Fig. 3C-F).

271 To further quantify how SNP rate differences between the strains biases reporting of differential
272 expression, we placed each of the 25,049 gene models with SNP rate data into one of seven "SNP rate
273 difference" categories for each pairwise strain comparison (data for the MHco3(ISE) vs. MHco4(WRS)
274 pairwise comparison is shown in Figure 4, and Supplementary Table S2). The percentage of gene

models reported as differentially expressed (with adjusted p-values < 0.05 and > log2 1X fold-change in expression) was lowest for the *0%* SNP rate difference category and increased as the SNP rate difference category increased (Fig. 4A). This trend was seen at all three SNP mapping allowances (Fig. 4A). There was also a strong relationship between the directionality of the differential expression called by *DESeq2* and the directionality of the SNP rate difference between the strains. For SNP rate difference categories where the SNP rate was greater in MHco4(WRS) than in MHco3(ISE) by at least 2%, the large majority of gene models reported as differentially expressed were down-regulated in MHco4(WRS) relative to MHco3(ISE) (396/425 (93.2%)) (Supplementary Table S2). Conversely, the large majority of gene models with SNP rates at least 2% lower in MHco4(WRS) than in MHco3(ISE), were up-regulated in MHco4(WRS) relative to MHco3(ISE) (21/27 (77.8%)) (Supplementary Table S2).

3.4 Minimizing the effect of sequence polymorphism differences on differential expression analysis in pairwise strain comparisons.

We next investigated ways to minimize the effect of sequence polymorphism on global transcriptomic differential expression analysis in pairwise strain comparisons. We first examined the effect of changing the read mapping polymorphism allowance on the number and bias of the differentially expressed genes reported by *DESeq2* in pairwise strain comparisons. When the polymorphism allowance was changed from N2 to N5 or from N5 to N10, there was an overall decrease in the total number of differentially expressed genes reported in all three pairwise strain comparisons (Supplementary Table S3). This trend was generally observed for genes in all SNP rate difference categories (see example of MHco3(ISE) vs. MHco4(WRS) pairwise comparison in Fig. 4A).

298 At the default N2 polymorphism allowance, *DESeq2* reported more genes down-regulated than up-
299 regulated in both MHco4(WRS) and MHco10(CAVR) when each was compared to MHco3(ISE)
300 (Supplementary Fig. S1; Supplementary Table S3). This bias was reduced as the mapping allowance
301 was increased to N5 and then N10 (Supplementary Fig. S1; Supplementary Table S3). In contrast, the
302 MHco4(WRS) and MHco10(CAVR) pairwise comparison showed a relatively equal ratio of down-
303 regulated and up-regulated gene numbers even at the default N2 polymorphism allowance
304 (Supplementary Fig. S1; Supplementary Table S3).
305 We then calculated the net (overall mean) differential expression (NDE) of all gene models in
306 each of the seven "SNP rate difference" categories for each of the pairwise strain comparisons to see if
307 there was an overall directional bias to the data (data for the MHco4(WRS) and MHco3(ISE) pairwise
308 strain comparison is shown in Fig. 4B). The NDE in the direction MHco4(WRS) > MHco3(ISE) was
309 greatest for those gene models in the *5 - 15%* MHco4(WRS) > MHco3(ISE) SNP rate difference
310 category and least for gene models in the 0% SNP rate difference category (Fig. 4B, Supplementary
311 Table S2A). Conversely, the NDE in the direction MHco4(WRS) < MHco3(ISE) was highest for gene
312 models in the *5 - 15%* MHco4(WRS) < MHco3(ISE) SNP rate difference category and least for the 0%
313 SNP rate difference category (Fig. 4B, Supplementary Table S2A). The NDE of gene models between
314 strains was highest at the N2 polymorphism mapping allowance, and least for the N10 polymorphism
315 mapping allowance, in all SNP rate difference categories (Fig. 4B; Supplementary Table S2A).
316 The NDE of gene models between the strains was relatively close to zero for genes of the three
317 lowest SNP rate difference categories, particularly at the N5 and N10 polymorphism allowances (Fig.
318 4B; Supplementary Table S2B). This suggests that gene models with < 2% difference in SNP rate
319 between strains had a minimal bias in pairwise strain differential expression analyses. We defined these
320 gene models as "low-polymorphic gene models" (LPGs) in the subsequent differential expression

analysis. These represent 17,881 out of the total of 25,111 gene models in the *H. contortus* whole genome annotation (71.2%) and so represent the majority of gene models (Supplementary Fig. S2).

3.5 Investigating genuine transcriptional differences between H. contortus strains.

We restricted the global transcriptomic analysis to the low-polymorphic gene models, as defined above, and used an N5 polymorphism allowance for read mapping to minimize the confounding effect of inter-strain sequence polymorphism. This resulted in the inclusion of 20,781, 19,397, and 22,924 gene models for the MHco4(WRS) vs. MHco3(ISE), MHco10(CAVR) vs. MHco3(ISE), and MHco4(WRS) vs. MHco10(CAVR) pairwise strain comparisons respectively (Supplementary Fig. S2). A set of 17,881 genes was common to the analysis set for all three pairwise comparisons (Supplementary Fig. S2). Normalized global expression of each of the nine bioreplicate RNAseq datasets clustered by strain on PCA analysis demonstrating that there are transcriptomic differences between the strains, even after the effects of sequence polymorphism on RNAseq mapping are minimized (Supplementary Fig. S3).

A total of 1,125 (5.41% of LPGs), 1,498 (7.72% of LPGs), and 824 (3.59% of LPGs) genes were differentially expressed at > 1X log2 fold in the MHco4(WRS) vs. MHco3(ISE), MHco10(CAVR) vs. MHco3(ISE), and MHco4(WRS) vs. MHco10(CAVR) pairwise comparisons respectively (Fig. 5). Of these, 134 genes (41 up-regulated, 93 down-regulated), 259 genes (121 up-regulated, 138 down regulated), and 103 genes (40 up-regulated, 63 down regulated) were > 2X log2 fold differentially expressed respectively (Fig. 5). The large majority of the most differentially expressed genes in all strains comparisons were either undescribed or had only broad ontological classifications (Supplementary Table S4). No previously reported ivermectin resistance candidate low-polymorphic

genes were observed to be differentially expressed in at > 2X log2 fold-change expression in either of the two ivermectin resistance strains relative to the MHCo3(ISE) susceptible strain (Supplementary Table S4).

We examined the number of genes that were differentially expressed in more than one of the pairwise strain comparisons to see if a set of genes was common to different pairwise comparisons. The highest proportion of shared differentially expressed LPGs was between the MHco4(WRS) vs. MHco3(ISE) and MHco10(CAVR) vs. MHco3(ISE) pairwise strain comparisons (Supplementary Fig. S4). Of the 2,132 gene models differentially expressed between either MHco4(WRS) and MHco10(CAVR) vs. MHco3(ISE), 491 (23.03%) were differentially expressed with the same directionality (up- or down- regulated) in both pairwise comparisons at >1X log2 fold change (48 gene models at > 2X log2 fold change) (Supplementary Fig. S4A). Fewer genes were shared in the other two strain combinations: of the 2,025 gene models differentially expressed between either MHco3(ISE) and MHco4(WRS) strains vs. MHco10(CAVR), 297 (14.67%) gene models were differentially expressed with the same directionality at >1 log2-fold change (39 gene models at >2 log2-fold change) in both pairwise comparisons (Supplementary Fig. S4B). Of the 1,794 gene models differentially expressed between either MHco3(ISE) and MHco10(CAVR) vs. MHco4(WRS), only 155 (8.64%) gene models were differentially expressed at >1 log2-fold change (8 gene models at >2 log2 fold change) with the same directionality in both comparisons (Supplementary Fig. S4C). Both these percentages represent a significantly lower proportion of differentially expressed genes shared than were observed shared in MHco4(WRS) and MHco10(CAVR) vs. MHco3(ISE) (Z-stats = 6.8 ($p < 0.000$), and 12.1 ($p < 0.000$) respectively).

366 *3.6 Investigating the effect of sequence polymorphism on differential expression analysis of two*
367 *gene families of relevance to ivermectin resistance research*.
368
369 67 ligand-gated chloride channels (LGICs) and 86 ABC transporters identified in the published *H.*
370 *contortus* draft genome (Laing et al, 2013) were examined for differential expression between the
371 MHco4(WRS) and MHco10(CAVR) ivermectin resistant strains and the susceptible MHco3(ISE)
372 strain. Three different differential expression analyses were compared to assess the impact of
373 accounting for sequence polymorphisms differences between the strains; using the default N2 SNP
374 allowance on all 25,111 gene models, using the N5 SNP allowance on all 25,111 genes, and using the
375 N5 SNP allowance on the set of 17,881 low-polymorphic genes (LPGs). There was a substantial
376 reduction in the total number of differentially expressed genes reported using the N5 allowance on the
377 LPG gene set compared with the N2 default allowance on the full gene set (Table 1). When comparing
378 the two ivermectin resistant strains with the ivermectin sensitive strain, only three of the low-
379 polymorphic genes – *Hco-lgc-55*, *Hco-pmp-6*, and *Hco-lgc-44* – showed differential expression at the
380 N5 allowance in both the MHco4(WRS) and MHco10(CAVR) vs. MHco3(ISE) pairwise comparisons.
381 *Hco-lgc-55* had > 2X log2 fold up-regulation in both cases (Table 1).
382
383 **4. Discussion**
384
385 Differential expression analysis, either at the single gene or whole transcriptome level, between
386 parasitic nematode strains and isolates is a common experimental approach. For example, a number of
387 candidate anthelmintic resistance genes have been identified by differential expression analysis of drug
388 resistant and susceptible isolates (Dicker et al., 2011; El-Abdellati et al., 2011; Williamson et al., 2011;

389 Xu et al., 1998). In the case of *H. contortus,* we reasoned that the extremely high levels of sequence
390 polymorphism both within and between laboratory strains and field isolates (reviewed in Gilleard and
391 Redman, (2016)), might confound the validity of such comparisons when using RNAseq, which is now
392 the central approach to conducting differential gene expression analyses. The majority of researchers
393 use only the default parameters of RNAseq data analysis pipelines and do not explore the effect of
394 different parameters on results reported (Baruzzo et al., 2017). It has been shown, using simulated
395 datasets, that the parameter with the greatest impact on performance is the number of mismatches
396 tolerated by during read mapping (Baruzzo et al., 2017). Since this seemed likely to be a particular
397 issue for organisms with high levels of sequence polymorphism, we undertook a detailed analysis to
398 examine the extent to which this may impact RNAseq based differential expression analysis between
399 *H. contortus* strains, and investigate how it could be mitigated to allow genuine transcriptional
400 differences to be assessed. We used *TopHat2* (Dobin and Gingeras, 2013) as our read mapping
401 software as this has been the mapping program most commonly used for RNAseq analysis over a
402 number of years and currently has the most citation in RNAseq literature. There are a number of
403 alternative mapping tools available whose use is becoming increasingly common, such as *HISAT2*
404 (Kim et al., 2015), which is *TopHat2's* recommended successor, but these tools are similarly sensitive
405 to changes in the mismatch parameter (Baruzzo et al., 2017).
406 A higher percentage of RNAseq reads mapped to the reference genome assembly for MHco3(ISE)
407 than for the MHco4(WRS) and MHco10(CAVR) strains (Fig. 1). This was hypothesized to be due to
408 sequence polymorphism reducing read mapping efficiency and reflecting the higher overall CDS SNP
409 rate in the latter two strains with respect to the MHco3(ISE) derived reference genome sequence (Fig.
410 1). This hypothesis was supported by the improvement of overall read mapping efficiency achieved by
411 increasing SNP mapping allowance to N5 (allowing 5 SNPs and 3 indels per read) from the default N2

412 value (allowing 2 SNPs or 2 indels per read). This change in SNP mapping allowance resulted in an
413 increase in the number of reads mapped for a large number of gene models (Fig. 2A). This
414 improvement in read mapping efficiency, as a result of increased SNP mapping allowance, was not
415 confined to the MHco4(WRS) and MHco10(CAVR) data, but also occurred with the MHco3(ISE) data.
416 These results suggest that mapping efficiency is affected by both between-strain and within-strain
417 sequence polymorphism. We also investigated the extent to which sequence polymorphism varied
418 among gene models and how this affected read mapping efficiency (Fig. 2B). When SNP allowances
419 were increased from N2 to N5, genes with higher levels of polymorphism showed larger proportionate
420 increases in reads mapped for all three strains (Fig. 2C, panel i). This further illustrates the impact of
421 sequence polymorphism on RNAseq read mapping efficiency and how it is greater for more
422 polymorphic genes.
423 Having shown that sequence polymorphism affects RNAseq read mapping to a reference genome
424 assembly with *TopHat2*, we next investigated how this might bias differential expression analysis using
425 *DESeq2*; one of the most commonly used bioinformatic tools for RNAseq data analysis (Fig. 3 and Fig.
426 4A). For each gene model, we plotted the *DESeq2* differential expression results against the difference
427 in SNP rate (relative to the reference genome assembly) between the two strains being compared (Fig.
428 3). For each pairwise strain comparison, gene models which had higher differences in the level of
429 sequence polymorphism between the strains were more likely to be down-regulated than to be up-
430 regulated in the strain with the highest level of sequence polymorphism (Fig. 3). Further, this bias
431 increased with the magnitude of difference in polymorphism rate of gene models between the strains
432 (Fig. 3 and Fig. 4A). This effect was true for all three pairwise strain comparisons, including between
433 the two “non-reference” MHco4(WRS) and MHco10(CAVR) strains. There is no obvious biological

reason for such differential expression biases, based on differences in SNP polymorphism rates, and so we concluded this is due to the effect of sequence polymorphism on RNAseq mapping rates.

Consequently, biases due to inter-strain differences in SNP polymorphism rates needed to be minimized before meaningful differential expression analysis could be performed. The first approach to achieve this was to choose RNAseq read mapping parameters in *TopHat2* to maximize read mapping efficiency for all the strains. Overall read mapping success peaked at the N5 or N6 SNP mapping allowances, depending on the strain (with very little difference between these two values (Fig. 1)). At the level of the gene model, the clear majority of genes had higher numbers of reads mapping at the N5 allowance than at either the N2 or N10 allowances (Fig. 2A). Consequently, the N5 mapping allowance maximized read mapping efficiency. Furthermore, the directional biases in the differential expression reports between strains were greatly reduced at the N5 mapping allowance (Fig. 4A-B, Supplementary Fig. S1). Consequently, the N5 mapping allowance was considered optimal to use for further analysis. However, optimizing the SNP mapping allowance did not completely remove the directional expression biases. For example, even at the N5 SNP mapping allowance, although the directional expression bias was close to zero for genes with SNP rate difference between strains of < 2%, it persisted for genes with a difference in SNP rate of > 2% (Fig. 4B). This led us to conclude that it was not possible to reliably measure differential expression for those genes > 2% SNP rate differences between strains, even at the N5 read mapping allowance. Consequently, we precluded these genes in subsequent transcriptomic analysis. These results have important implications for differential expression analysis between different strains/isolates of organisms with high levels of genetic diversity and suggest that sequence polymorphism needs to be defined and accounted for as part of the analysis. There are an number of other read mapping tools available for RNAseq analysis some of which, although less widely used than *TopHat2*, may be less impacted by high levels of sequence

polymorphism (Baruzzo et al., 2017). *TopHat2* is still widely used but it is noteworthy that the mapping tool which is increasingly used in place *TopHat2* is *HISAT2*, which is only slightly less sensitive to changes in mismatch parameters using simulated datasets (Baruzzo et al., 2017). Other read mapping tools such as *NovoAlign* (http://www.novocraft.com/products/novoalign/) or *GSNAP* (Wu and Nacu, 2010), that may be less impacted by sequence polymorphism, deserve more exploration for use in RNAseq differential expression pipelines for organisms such as *H. contortus* with high levels of genetic variation.

Pairwise comparisons of three genetically divergent strains of *H. contortus* revealed large numbers of differentially expressed genes, even after the confounding effects of sequence polymorphism were removed (Fig. 5). The proportion of differentially expressed genes between the *H. contortus* strains far exceed those previously observed in inter-population studies of vertebrate species such as human and mouse (Bottomly et al., 2011; Li et al., 2014), and it is greater than has been reported between different strains of *C. elegans* (N2/Bristol and CB4856/Hawaiian strains) (Capra et al., 2008; Francesconi and Lehner, 2014).This remarkably large number of differentially expressed genes between these *H. contortus* strains may have many different phenotypic traits which could have a variety of implications for their life history traits, epidemiology, pathogenicity, and susceptibility to drugs and/or vaccines. This reflects the high genetic diversity of *H. contortus* and of these particular strains. MHco3(ISE), MHco4(WRS), and MHco10(CAVR) are derived from field isolates obtained from different continents and are highly genetically divergent (Gilleard and Redman, 2016; Redman et al., 2012, 2008). For example, the levels of genetic diversity (Fst values) between strains based on microsatellite genotyping ranged from 0.1530 to 0.2696 which is as high or higher than some closely related species in some cases (Prado-Martinez et al., 2013; Redman et al., 2008; Romiguier et al., 2014). Further, although the nematode body plan is superficially simple, a variety of morphological and morphometric traits vary

between these three strains, including vulval morphology, oesophagus length, and spicule length in males as well as the extent of the synlophe cuticular ridges in females (Gilleard and Redman, 2016; Sargison et al., 2019). Also, there is evidence of lethality of some hybrid progeny of these strains (Sargison et al., 2019).

The results of this study also have important implications for anthelmintic resistance research which, until very recently, has been dominated by candidate gene studies (Gilleard, 2013, 2006; Rezansoff et al., 2016). In the case of ivermectin resistance, such studies have so far failed to identify the key loci or genes involved in resistance for any parasitic nematode, including *H. contortus* (Gilleard, 2013). One common component of candidate gene studies has been to compare the expression levels of specific candidate genes between a small number of ivermectin resistant and susceptible parasite strains (Dicker et al., 2011; El-Abdellati et al., 2011; Williamson et al., 2011; Xu et al., 1998). It is common for such studies to report differences in expression between resistant and susceptible strains for candidate genes such as P-glycoproteins (PGPs) or ligand-gated ion channels (LGICs). These differences are commonly used as circumstantial evidence for a role in resistance. Our results here show the context in which such studies should be interpreted as a very large number of genes are differentially expressed in pairwise comparisons of genetically divergent *H. contortus* strains (Fig. 5). 824 - 1,498 low-polymorphic genes were differentially expressed between the strains in the study at a level of 2-fold and an adjusted statistical significance of $p < 0.05$ (as called by *DESeq2*). This highlights the inherently high levels of "background" transcriptomic variation that occur between genetically divergent *H. contortus* strains. Consequently, care must be taken when interpreting a suggested association of differential expression of a gene with a drug resistance phenotype when a small number of genes are compared between a small number of drug resistant and susceptible strains.

502 This is particularly the case when the degree of genetic differentiation or the general level of
503 transcriptomic difference that exists between the strains has not been assessed.
504 Recently, studies analyzing the expression of small numbers of candidate genes are being replaced
505 with more global transcriptomic studies. The draft *H. contortus* genome and its recent improvement
506 into a chromosomal level assembly is making such studies increasingly feasible on a genome-wide
507 scale (Doyle et al., 2018; Laing et al., 2013). The work presented here also has important implications
508 for global transcriptomic comparisons of drug resistant and susceptible strains. Two gene families often
509 suggested to be involved in ivermectin resistance are the LGICs and ABC transporter genes (Laing et
510 al., 2013). We used the gene models in the *H. contortus* draft annotation to assess how many members
511 of these gene families were differentially expressed between the MHco4(WRS) and MHco10(CAVR)
512 ivermectin resistant strains and the MHco3(ISE) susceptible strain using the default polymorphism
513 allowance (N2), the optimized polymorphism allowance (N5), and the polymorphism allowance (N5)
514 but removing the highly polymorphic gene set (Table 1). We found there was a dramatic reduction in
515 the number of members of these genes families that were determined to be differentially expressed
516 when polymorphism allowance was increased to the optimal N5 allowance (Table 1). A further
517 reduction was apparent when the most highly polymorphic genes were discarded from the analysis
518 (Table 1).
519 These results highlight the fact that a substantial number of differentially expressed genes reported
520 are likely to be artifacts caused by differences in sequence polymorphism between the strains being
521 compared which are not accounted for. In the case of our analysis, accounting for sequence
522 polymorphism reveals a smaller number of differentially expressed candidate genes perhaps worthy of
523 further investigation. The ABC transporter *Hco-pmp-6*, and two LGICs – *Hco-lgc-55* and *Hco-lgc-44* –
524 were differentially expressed with the same directionality in both ivermectin resistant strains relative to

the MHco3(ISE) strain. *Hco-lgc-55* is a tyramine-gated chloride channel whose *C*. *elegans* homologue *Cel-lgc-55* is expressed in the pharynx and is involved in worm motility (Rao et al., 2010; Ringstad et al., 2009). The ABC transporter *Hco-wht-4*, and the LGICs *Hco-lgc-3*, *Hco-lgc-33*, *Hco-lgc-9*, and *Hco-acr-24* were other genes with a > 2X log2 fold-change differential expression in the MHco10(CAVR) strain, although these genes were not differentially expressed in the other resistant strain, MHco4(WRS). *Hco-lgc-3* was the gene with the highest level of up-regulation across both these gene families, being differentially expressed at greater than 50-fold in MHco3(CAVR) relative to MHco3(ISE) (Table 1). The gene may be considered of interest given its homology to a paralogous pair of *C*. *elegans* proton-gated ion channels, *Cel-pbo-5* and *Cel-pbo-6,* which are required for normal posterior muscle function (Beg et al., 2008). However, further functional and genetic studies are required before making any inferences of the potential role of these genes in mediating the ivermectin resistance phenotype of *H*. *contortus*.

Acknowledgements

We are grateful to Dr Matt Workentine for comments on the manuscript. We are grateful for funding from NSERC Discovery Grant, NSERC-CREATE Host-Parasites Interactions (HPI) program and the Biotechnology and Biological Sciences Research Council (BBSRC). The Moredun Research Institute (DJB and AAM) receives funding from the Scottish Government. NS was funded from the Higher Education Funding Council of England (HEFCE), the Department for Environment, Food and Rural Affairs (DEFRA) and the Scottish Funding Council (SFC) Veterinary Training Research Initiative (VTRI) programme VT0102 and for the support of Pfizer Animal Health.

References

Anders, S., Pyl, P.T., Huber, W., 2014. HTSeq – A Python framework to work with high-throughput sequencing data HTSeq – A Python framework to work with high-throughput sequencing data. Bioinformatics. 31, 0–5. https://doi.org/10.1093/bioinformatics/btu638

Andersen, E.C., Gerke, J.P., Shapiro, J.A., Crissman, J.R., Ghosh, R., Bloom, J.S., Felix, M.-A., Kruglyak, L., 2012. Chromosome-scale selective sweeps shape Caenorhabditis elegans genomic

553 diversity. Nat. Genet. 44, 285–295. https://doi.org/10.1038/ng.1050
554 Antony, H.A., Pathak, V., Parija, S.C., Ghosh, K., Bhattacherjee, A., 2016. Transcriptomic Analysis of
555 Chloroquine-Sensitive and Chloroquine-Resistant Strains of Plasmodium falciparum : Toward
556 Malaria Diagnostics and Therapeutics for Global Health. Omi. A J. Integr. Biol. 20, 424–432.
557 https://doi.org/10.1089/omi.2016.0058
558 Baruzzo, G., Hayer, K.E., Kim, E.J., DI Camillo, B., Fitzgerald, G.A., Grant, G.R., 2017. Simulation-
559 based comprehensive benchmarking of RNA-seq aligners. Nat. Methods 14, 135–139.
560 https://doi.org/10.1038/nmeth.4106
561 Bateman, A., Martin, M.J., O'Donovan, C., Magrane, M., Apweiler, R., Alpi, E., Antunes, R.,
562 Arganiska, J., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Chavali, G., Cibrian-
563 Uhalte, E., Da Silva, A., De Giorgi, M., Dogan, T., Fazzini, F., Gane, P., Castro, L.G., Garmiri, P.,
564 Hatton-Ellis, E., Hieta, R., Huntley, R., Legge, D., Liu, W., Luo, J., Macdougall, A., Mutowo, P.,
565 Nightingale, A., Orchard, S., Pichler, K., Poggioli, D., Pundir, S., Pureza, L., Qi, G., Rosanoff, S.,
566 Saidi, R., Sawford, T., Shypitsyna, A., Turner, E., Volynkin, V., Wardell, T., Watkins, X., Zellner,
567 H., Cowley, A., Figueira, L., Li, W., McWilliam, H., Lopez, R., Xenarios, I., Bougueleret, L.,
568 Bridge, A., Poux, S., Redaschi, N., Aimo, L., Argoud-Puy, G., Auchincloss, A., Axelsen, K.,
569 Bansal, P., Baratin, D., Blatter, M.C., Boeckmann, B., Bolleman, J., Boutet, E., Breuza, L., Casal-
570 Casas, C., De Castro, E., Coudert, E., Cuche, B., Doche, M., Dornevil, D., Duvaud, S., Estreicher,
571 A., Famiglietti, L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-
572 Gumowski, N., Hinz, U., Hulo, C., Jungo, F., Keller, G., Lara, V., Lemercier, P., Lieberherr, D.,
573 Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T., Nouspikel, N., Paesano, S.,
574 Pedruzzi, I., Pilbout, S., Pozzato, M., Pruess, M., Rivoire, C., Roechert, B., Schneider, M., Sigrist,
575 C., Sonesson, K., Staehli, S., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Veuthey, A.L.,
576 Wu, C.H., Arighi, C.N., Arminski, L., Chen, C., Chen, Y., Garavelli, J.S., Huang, H., Laiho, K.,
577 McGarvey, P., Natale, D.A., Suzek, B.E., Vinayaka, C.R., Wang, Q., Wang, Y., Yeh, L.S.,
578 Yerramalla, M.S., Zhang, J., 2015. UniProt: A hub for protein information. Nucleic Acids Res. 43,
579 D204–D212. https://doi.org/10.1093/nar/gku989
580 Beg, A.A., Ernstrom, G.G., Nix, P., Davis, M.W., Jorgensen, E.M., 2008. Protons Act as a Transmitter
581 for Muscle Contraction in C. elegans. Cell 132, 149–160.
582 https://doi.org/10.1016/j.cell.2007.10.058
583 Blumenthal, T., Davis, R.E., 2004. Exploring nematode diversity. Nat. Genet.
584 https://doi.org/10.1038/ng1204-1246
585 Bottomly, D., Walter, N.A.R., Hunter, J.E., Darakjian, P., Kawane, S., Buck, K.J., Searles, R.P.,
586 Mooney, M., McWeeney, S.K., Hitzemann, R., 2011. Evaluating gene expression in C57BL/6J
587 and DBA/2J mouse striatum using RNA-Seq and microarrays. PLoS One 6.
588 https://doi.org/10.1371/journal.pone.0017820
589 Capra, E.J., Skrovanek, S.M., Kruglyak, L., 2008. Comparative developmental expression profiling of
590 two C. elegans isolates. PLoS One 3. https://doi.org/10.1371/journal.pone.0004055
591 Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden,
592 D.M., 2012. A program for annotating and predicting the effects of single nucleotide
593 polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain. Fly (Austin). 6,
594 80–92. https://doi.org/10.4161/fly.19695
595 Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak,
596 M.W., Gaffney, D.J., Elo, L.L., Zhang, X., Mortazavi, A., 2016. A survey of best practices for
597 RNA-seq data analysis. Genome Biol. https://doi.org/10.1186/s13059-016-0881-8
598 Croken, M.M.K., Ma, Y., Markillie, L.M., Taylor, R.C., Orr, G., Weiss, L.M., Kim, K., 2014. Distinct

599 strains of Toxoplasma gondii feature divergent transcriptomes regardless of developmental stage.
600 PLoS One 9, 1–10. https://doi.org/10.1371/journal.pone.0111297
601 Dey, A., Chan, C.K.W., Thomas, C.G., Cutter, A.D., 2013. Molecular hyperdiversity defines
602 populations of the nematode Caenorhabditis brenneri. Proc. Natl. Acad. Sci. 110, 11056–11060.
603 https://doi.org/10.1073/pnas.1303057110
604 Dicker, A.J., Nisbet, A.J., Skuce, P.J., 2011. Gene expression changes in a P-glycoprotein (Tci-pgp-9)
605 putatively associated with ivermectin resistance in Teladorsagia circumcincta. Int. J. Parasitol. 41,
606 935–942. https://doi.org/10.1016/j.ijpara.2011.03.015
607 Dobin, A., Gingeras, T.R., 2013. Comment on "TopHat2: accurate alignment of transcriptomes in the
608 presence of insertions, deletions and gene fusions" by Kim et al. 2013. bioRxiv 0–9.
609 https://doi.org/10.1101/000851
610 Doyle, S.R., Illingworth, C.J.R., Laing, R., Bartley, D.J., Redman, E., Martinelli, A., Holroyd, N.,
611 Morrison, A.A., Rezansoff, A., Tracey, A., Devaney, E., Berriman, M., Sargison, N., Cotton, J.A.,
612 Gilleard, J.S., 2019. Population genomic and evolutionary modelling analyses reveal a single
613 major QTL for ivermectin drug resistance in the pathogenic nematode, Haemonchus contortus.
614 BMC Genomics 20, 218. https://doi.org/10.1186/s12864-019-5592-6
615 Doyle, S.R., Laing, R., Bartley, D.J., Britton, C., Chaudhry, U., Gilleard, J.S., Holroyd, N., Mable,
616 B.K., Maitland, K., Morrison, A.A., Tait, A., Tracey, A., Berriman, M., Devaney, E., Cotton, J.A.,
617 Sargison, N.D., 2018. A Genome Resequencing-Based Genetic Map Reveals the Recombination
618 Landscape of an Outbred Parasitic Nematode in the Presence of Polyploidy and Polyandry.
619 Genome Biol. Evol. 10, 396–409. https://doi.org/10.1093/gbe/evx269
620 Edwards, J.A., Chen, C., Kemski, M.M., Hu, J., Mitchell, T.K., Rappleye, C.A., 2013. Histoplasma
621 yeast and mycelial transcriptomes reveal pathogenic-phase and lineage-specific gene expression
622 profiles. BMC Genomics 14, 695. https://doi.org/10.1186/1471-2164-14-695
623 El-Abdellati, A., De Graef, J., Van Zeveren, A., Donnan, A., Skuce, P., Walsh, T., Wolstenholme, A.,
624 Tait, A., Vercruysse, J., Claerebout, E., Geldhof, P., 2011. Altered avr-14B gene transcription
625 patterns in ivermectin-resistant isolates of the cattle parasites, Cooperia oncophora and Ostertagia
626 ostertagi. Int. J. Parasitol. 41, 951–957. https://doi.org/10.1016/j.ijpara.2011.04.003
627 Fiebig, M., Kelly, S., Gluenz, E., 2015. Comparative Life Cycle Transcriptomics Revises Leishmania
628 mexicana Genome Annotation and Links a Chromosome Duplication with Parasitism of
629 Vertebrates. PLoS Pathog. 11, 1–28. https://doi.org/10.1371/journal.ppat.1005186
630 Francesconi, M., Lehner, B., 2014. The effects of genetic variation on gene expression dynamics
631 during development. Nature 505, 208–211. https://doi.org/10.1038/nature12772
632 Gilleard, J.S., 2013. Haemonchus contortus as a paradigm and model to study anthelmintic drug
633 resistance. Parasitology 140, 1506–1522. https://doi.org/10.1017/S0031182013001145
634 Gilleard, J.S., 2006. Understanding anthelmintic resistance: The need for genomics and genetics. Int. J.
635 Parasitol. https://doi.org/10.1016/j.ijpara.2006.06.010
636 Gilleard, J.S., Redman, E., 2016. Genetic Diversity and Population Structure of Haemonchus contortus,
637 in: Advances in Parasitology. pp. 31–68. https://doi.org/10.1016/bs.apar.2016.02.009
638 Guryev, V., Koudijs, M.J., Berezikov, E., Johnson, S.L., Plasterk, R.H.A., van Eeden, F.J.M., Cuppen,
639 E., 2006. Genetic variation in the zebrafish. Genome Res. 16, 491–7.
640 https://doi.org/10.1101/gr.4791006
641 Kim, D., Langmead, B., Salzberg, S.L., 2015. HISAT: a fast spliced aligner with low memory
642 requirements. Nat. Methods 12, 357–360. https://doi.org/10.1038/nmeth.3317
643 Laing, R., Hunt, M., Protasio, A. V, Saunders, G., Mungall, K., Laing, S., Jackson, F., Quail, M.,
644 Beech, R., Berriman, M., Gilleard, J.S., 2011. Annotation of two large contiguous regions from

the Haemonchus contortus genome using RNA-seq and comparative analysis with Caenorhabditis elegans. PLoS One 6, e23216. https://doi.org/10.1371/journal.pone.0023216

Laing, R., Kikuchi, T., Martinelli, A., Tsai, I., Beech, R., Redman, E., Holroyd, N., Bartley, D., Beasley, H., Britton, C., Curran, D., Devaney, E., Gilabert, A., Hunt, M., Jackson, F., Johnston, S., Kryukov, I., Li, K., Morrison, A., Reid, A., Sargison, N., Saunders, G., Wasmuth, J., Wolstenholme, A., Berriman, M., Gilleard, J., Cotton, J., 2013. The genome and transcriptome of Haemonchus contortus, a key model parasite for drug and vaccine discovery. Genome Biol. 14, R88. https://doi.org/10.1186/gb-2013-14-8-r88

Le Jambre, L., Gill, J., Lenane, I., Lacey, E., 1995. Characterization of an avermectin resistant strain of autralian Haemonchus contortus. Int. J. Parasitol. 25, 691–698.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Subgroup, 1000 Genome Project Data Processing, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, J., Lai, K., Ching, A.K.K., Chan, T., 2014. Genomics Transcriptome sequencing of Chinese and Caucasian population identi fi es ethnic-associated differential transcript abundance of heterogeneous nuclear ribonucleoprotein K (hnRNPK). Genomics 103, 56–64. https://doi.org/10.1016/j.ygeno.2013.12.005

Lindblad-Toh, K., Lander, E.S., Winchester, E., Daly, M.J., Wang, D.G., Hirschhorn, J.N., Laviolette, J.-P., Ardlie, K., Reich, D.E., Robinson, E., Sklar, P., Shah, N., Thomas, D., Fan, J.-B., Gingeras, T., Warrington, J., Patil, N., Hudson, T.J., 2000. Large-scale discovery and genotyping of single-nucleotide polymorphismsin the mouse. Nat. Genet. 24, 381–386. https://doi.org/10.1038/74215

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 1–21. https://doi.org/10.1186/s13059-014-0550-8

Papenfort, K., Förstner, K.U., Cong, J., Sharma, C.M., Bassler, B.L., 2015. Differential RNA-seq of Vibrio cholerae identifies the VqmR small RNA as a regulator of biofilm formation. Proc. Natl. Acad. Sci. 112, E766–E775. https://doi.org/10.1073/pnas.1500203112

Pathan, M., Keerthikumar, S., Ang, C.S., Gangoda, L., Quek, C.Y.J., Williamson, N.A., Mouradov, D., Sieber, O.M., Simpson, R.J., Salim, A., Bacic, A., Hill, A.F., Stroud, D.A., Ryan, M.T., Agbinya, J.I., Mariadason, J.M., Burgess, A.W., Mathivanan, S., 2015. FunRich: An open access standalone functional enrichment and interaction network analysis tool. Proteomics 15, 2597–2601. https://doi.org/10.1002/pmic.201400515

Prado-Martinez, J., Sudmant, P.H., Kidd, J.M., Li, H., Kelley, J.L., Lorente-Galdos, B., Veeramah, K.R., Woerner, A.E., O'Connor, T.D., Santpere, G., Cagan, A., Theunert, C., Casals, F., Laayouni, H., Munch, K., Hobolth, A., Halager, A.E., Malig, M., Hernandez-Rodriguez, J., Hernando-Herraez, I., Prüfer, K., Pybus, M., Johnstone, L., Lachmann, M., Alkan, C., Twigg, D., Petit, N., Baker, C., Hormozdiari, F., Fernandez-Callejo, M., Dabad, M., Wilson, M.L., Stevison, L., Camprubí, C., Carvalho, T., Ruiz-Herrera, A., Vives, L., Mele, M., Abello, T., Kondova, I., Bontrop, R.E., Pusey, A., Lankester, F., Kiyang, J.A., Bergl, R.A., Lonsdorf, E., Myers, S., Ventura, M., Gagneux, P., Comas, D., Siegismund, H., Blanc, J., Agueda-Calpena, L., Gut, M., Fulton, L., Tishkoff, S.A., Mullikin, J.C., Wilson, R.K., Gut, I.G., Gonder, M.K., Ryder, O.A., Hahn, B.H., Navarro, A., Akey, J.M., Bertranpetit, J., Reich, D., Mailund, T., Schierup, M.H., Hvilsom, C., Andrés, A.M., Wall, J.D., Bustamante, C.D., Hammer, M.F., Eichler, E.E., Marques-Bonet, T., 2013. Great ape genetic diversity and population history. Nature 499, 471–475. https://doi.org/10.1038/nature12228

Rao, V.T.S., Accardi, M. V., Siddiqui, S.Z., Beech, R.N., Prichard, R.K., Forrester, S.G., 2010. Characterization of a novel tyramine-gated chloride channel from Haemonchus contortus. Mol.

691 Biochem. Parasitol. 173, 64–68. https://doi.org/10.1016/j.molbiopara.2010.05.005
692 Redman, E., Packard, E., Grillo, V., Smith, J., Jackson, F., Gilleard, J.S., 2008. Microsatellite analysis
693 reveals marked genetic differentiation between Haemonchus contortus laboratory isolates and
694 provides a rapid system of genetic fingerprinting. Int. J. Parasitol. 38, 111–22.
695 https://doi.org/10.1016/j.ijpara.2007.06.008
696 Redman, E., Sargison, N., Whitelaw, F., Jackson, F., Morrison, A., Bartley, D.J., Gilleard, J.S., 2012.
697 Introgression of Ivermectin Resistance Genes into a Susceptible Haemonchus contortus Strain by
698 Multiple Backcrossing. PLoS Pathog. 8, e1002534. https://doi.org/10.1371/journal.ppat.1002534
699 Redman, E., Whitelaw, F., Tait, A., Burgess, C., Bartley, Y., Skuce, P.J., Jackson, F., Gilleard, J.S.,
700 2015. The emergence of resistance to the benzimidazole anthlemintics in parasitic nematodes of
701 livestock is characterised by multiple independent hard and soft selective sweeps. PLoS Negl.
702 Trop. Dis. 9, e0003494. https://doi.org/10.1371/journal.pntd.0003494
703 Rezansoff, A.M., Laing, R., Gilleard, J.S., 2016. Evidence from two independent backcross
704 experiments supports genetic linkage of microsatellite Hcms8a20, but not other candidate loci, to
705 a major ivermectin resistance locus in Haemonchus contortus. Int. J. Parasitol. 46, 653–661.
706 https://doi.org/10.1016/j.ijpara.2016.04.007
707 Ringstad, N., Abe, N., Horvitz, H.R., 2009. Ligand-gated chloride channels are receptors for biogenic
708 amines in C. elegans. Science 325, 96–100. https://doi.org/10.1126/science.1169243
709 Romiguier, J., Gayral, P., Ballenghien, M., Bernard, A., Cahais, V., Chenuil, A., Chiari, Y., Dernat, R.,
710 Duret, L., Faivre, N., Loire, E., Lourenco, J.M., Nabholz, B., Roux, C., Tsagkogeorga, G., Weber,
711 A.A.T., Weinert, L.A., Belkhir, K., Bierne, N., Glémin, S., Galtier, N., 2014. Comparative
712 population genomics in animals uncovers the determinants of genetic diversity. Nature 515, 261–
713 263. https://doi.org/10.1038/nature13685
714 Sargison, N.D., Redman, E., Morrison, A.A., Bartley, D.J., Jackson, F., Hoberg, E., Gilleard, J.S.,
715 2019. Mating barriers between genetically divergent strains of the parasitic nematode
716 Haemonchus contortus suggest incipient speciation. Int. J. Parasitol.
717 https://doi.org/10.1016/j.ijpara.2019.02.008
718 Urdaneta-Marquez, L., Bae, S.H., Janukavicius, P., Beech, R., Dent, J., Prichard, R., 2014. A dyf-7
719 haplotype causes sensory neuron defects and is associated with macrocyclic lactone resistance
720 worldwide in the nematode parasite Haemonchus contortus. Int. J. Parasitol. 44, 1063–1071.
721 https://doi.org/10.1016/j.ijpara.2014.08.005
722 Van Wyk, J.A., Malan, F.S., 1988. Resistance of field strains of Haemonchus contortus to ivermectin,
723 closantel, rafoxanide and the benzimidazoles in South Africa. Vet. Rec. 123, 226–228.
724 https://doi.org/10.1136/vr.123.9.226
725 Wang, D.G., 1998. Large-Scale Identification, Mapping, and Genotyping of Single-Nucleotide
726 Polymorphisms in the Human Genome. Science (80-.). 280, 1077–1082.
727 https://doi.org/10.1126/science.280.5366.1077
728 Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat.
729 Rev. Genet. 10, 57–63. https://doi.org/10.1038/nrg2484
730 Williamson, S.M., Storey, B., Howell, S., Harper, K.M., Kaplan, R.M., Wolstenholme, A.J., 2011.
731 Candidate anthelmintic resistance-associated gene expression and sequence polymorphisms in a
732 triple-resistant field isolate of Haemonchus contortus. Mol. Biochem. Parasitol. 180, 99–105.
733 https://doi.org/10.1016/j.molbiopara.2011.09.003
734 Wu, T.D., Nacu, S., 2010. Fast and SNP-tolerant detection of complex variants and splicing in short
735 reads. Bioinformatics 26, 873–881. https://doi.org/10.1093/bioinformatics/btq057
736 Xu, M., Molento, M., Blackhall, W., Ribeiro, P., Beech, R., Prichard, R., 1998. Ivermectin resistance in

nematodes may be caused by alteration of P-glycoprotein homolog. Mol. Biochem. Parasitol. 91, 327–335. https://doi.org/10.1016/S0166-6851(97)00215-6

Figure Legends

Figure 1. The percentage of RNAseq reads that mapped to the MHco3(ISE) reference genome assembly at different *TopHat2* SNP (polymorphism) allowances (N2 to N10) shown for each of the three *H. contortus* strains MHco3(ISE), MHco4(WRS), and MHco10(CAVR).

Figure 2. A) The number of genes which had either a >1% increase (green bars) or >1% decrease (red bars) in the number of RNAseq reads mapping to them on the reference MHco3(ISE) genome assembly following an increase in the read mapping polymorphism allowance in *TopHat2* for *H. contortus* strains MHco3(ISE), MHco4(WRS), and MHco10(CAVR). Panel *i* shows the data for a change in polymorphism allowance of N2 to N5 and panel *ii* shows the data for a change from N5 to N10. B) The number of gene models in each SNP rate category for each *H. contortus* strain. The SNP rate for each gene model was calculated by dividing the number of SNPs in each CDS by the total CDS length for each gene model. C) Ratios of the total number of RNAseq reads mapping to gene models in each SNP rate category at two different SNP mapping allowances for each *H. contortus* strain. Panel *i* shows the N5:N2 ratio. Panel *ii* shows the N10:N5 ratio. Counts of reads mapped were totaled for all genes within each SNP rate category of each strain (colour coded).

Figure 3. Scatter plots of the differential expression of gene models, as determined by *DESeq2* (X-axis), plotted against their difference in SNP rate percentage between the two strains being compared (Y-axis). Gene model data points in each pairwise comparison are split into two panels, the left panel showing the gene models with higher SNP rates in one strain of each pairwise comparison and the right panel showing the gene models with higher SNP rates in the other pairwise strain. A and B show the MHco4(WRS) vs. MHco3(ISE) comparison, C and D show the MHco10(CAVR) vs. MHco3(ISE) comparison, and E and F show the MHco4(WRS) vs. MHco10(CAVR) comparison. The difference in the SNP rate percentage between the two strains is shown on the y-axis and plotted against reported log2 fold-change differential expression for each gene. The red lines represent zero differential expression.

Figure 4. A) The percentage of expressed gene models in each SNP rate difference category that are differentially expressed between MHco3(ISE) and MHco4(WRS) (log2 fold-change > 1X; adjusted p-value < 0.05) for each of the three SNP (polymorphism) allowances – N2, N5, and N10 – when mapping. B) The net log2 fold differences in expression (NDE) of all expressed genes in each SNP rate difference category. NDEs are shown for the N2, N5 and N10 SNP allowances when read mapping for the MHco3(ISE) vs. MHco4(WRS) pairwise comparison. NDEs are the mean value for all genes in each SNP rate difference category. Negative NDE values indicate an overall bias towards down-

regulation of genes in MHco4(WRS) vs. MHco3(ISE) strain. Positive values report an overall bias towards up-regulation of genes.

Figure 5. The total number of differentially expressed low-polymorphic genes (LPGs) observed in each pairwise strain comparison at the N5 mapping allowance. Gene counts at both > 1X log2 fold-change (orange dots), and > 2X log2 fold-change (red dots) thresholds are shown. The blue line on the y-axis represents an adjusted p-value of 0.05.

Supplementary Figure Legends

Supplementary Figure S1. Volcano plots showing differential expression of gene models at three different SNP allowances in *Tophap2's* mapping parameters (N2, N5, N10) are shown for each pairwise strain comparison. Log2 fold-change difference in expression from -4 to 4 is represented along the x-axis of each chart, and *DESeq2* -log10 adjusted p-values of the differential expression calls from 0 to 30 are represented along the y-axis. Gene positions exceeding a maximum value on either axis are placed at max value on that axis. Red points on the right and left sides of each plot represent genes differentially expressed at > 1X and < -1X log2 fold-change respectively with adjusted p-values < 0.05. Blue points represent genes significantly differentially expressed but at less than 1X log2 fold-change in either direction.

Supplementary Figure S2. Venn diagram showing the numbers of gene models qualifying as low-polymorphic genes to be included in the different pairwise strain comparisons. The total number of genes qualifying as low-polymorphic genes in each of the pairwise strain comparisons are shown outside respective circles (i.e. gene models with differences in SNP rates between the two strains of < 2%). The number of these genes shared and not shared among the pairwise strain comparisons are shown within respective Venn circles.

Supplementary Figure S3. A PCA plot representing the variance in log gene expression of low-polymorphic genes of each triplicate dataset for each of the three populations when mapped at the N5 mapping allowance.

Supplementary Figure S4. Venn diagrams showing the numbers of genes differentially expressed in each pairwise strain comparison, and shared differentially expressed between different pairwise strain comparisons. Venn circles are colour coded by pairwise strain comparison – red represents differentially expressed gene numbers of the MHco4(WRS) vs. MHco3(ISE) comparison, orange represents the MHco10(CAVR) vs. MHco3(ISE) comparison, and green represents the MHco4(WRS) vs. MHco10(CAVR) comparison. Differentially expressed genes were counted and cross-referenced at

two thresholds of differential expression: log2 fold-change difference in expression > 1 (italicized), and log2 fold-change difference in expression > 2 (bolded).

Tables.

Table 1. Differentially expressed genes of ivermectin resistance candidate gene families, the ligand-gated ion channels (LGICs) (A and B), and ABC transporters (C and D). The respective log2 fold-change differences in expression are shown for both of the ivermectin resistant strains – MHco4(WRS) (A and C) and MHco10(CAVR) (B and D) relative to the ivermectin susceptible reference strain – MHco3(ISE). In each panel, all genes differentially expressed at the default (N2) SNP allowance are shown (left column), at the N5 SNP allowance (middle column), and at the N5 SNP allowance with the highly polymorphic genes removed (right column). Highly polymorphic genes comprise genes with a > 2% difference in SNP rate between the two strains compared.

Ligand-gated Ion Channels

Up-regulated in MHco4(WRS)

N2		N5		N5 LPGs	
Hco-lgc-55	2.64	*Hco-lgc-55*	2.74	*Hco-lgc-55*	2.74
Hco-acc-2	1.15				
Hco-lgc-39	1.11				

Up-regulated in MHco10(CAVR)

N2		N5		N5 LPGs	
Hco-lgc-3	6.83	*Hco-lgc-3*	5.87	*Hco-lgc-3*	5.87
Hco-lgc-55	2.36	*Hco-lgc-55*	2.33	*Hco-lgc-55*	2.33
Hco-acr-6	1.73	*Hco-acr-6*	1.53	*Hco-acr-6*	1.53
Hco-lgc-41	1.58	*Hco-lgc-41*	1.48	*Hco-lgc-41*	1.48
Hco-acc-1	1.53				
Hco-mptl-1	1.32	*mtpl-1*	1.32	*Hco-mptl-1*	1.32
Hco-glc-2	1.06				

Down-regulated in MHco4(WRS)

N2		N5		N5 LPGs	
Hco-glc-5	-2.56	*Hco-glc-5*	-1.47		
Hco-lgc-34	-2.41	*Hco-lgc-34*	-1.19	*Hco-lgc-34*	-1.19
Hco-lgc-7	-2.3	*Hco-lgc-7*	-2.55		
Hco-acr-17b	-1.78				
Hco-lgc-43	-1.61				
Hco-lgc-44	-1.5	*Hco-lgc-44*	-1.14	*Hco-lgc-44*	-1.14
Hco-acr-24	-1.29	*Hco-acr-24*	-1.44		
Hco-des-2	-1.18	*Hco-des-2*	-1.27	*Hco-des-2*	-1.27
Hco-lgc-27	-1.12				
Hco-lgc-40	-1.07				

Down-regulated in MHco10(CAVR)

N2		N5		N5 LPGs	
Hco-lgc-7	-4.48	*Hco-lgc-7*	-4.3		
Hco-glc-5	-3.02	*Hco-glc-5*	-1.88		
Hco-lgc-33	-2.99	*Hco-lgc-33*	-2.91	*Hco-lgc-33*	-2.91
Hco-lgc-43	-2.66	*Hco-lgc-43*	-1.37		
Hco-lgc-9	-2.26	*Hco-lgc-9*	-2.15	*Hco-lgc-9*	-2.15
Hco-acr-24	-2.09	*Hco-acr-24*	-2.17	*Hco-acr-24*	-2.17
Hco-ggr-3	-2.09	*Hco-ggr-3*	-1.69	*Hco-ggr-3*	-1.69
Hco-lgc-44	-1.5	*Hco-lgc-44*	-1.93	*Hco-lgc-44*	-1.93
Hco-eat-2X	-1.8				
Hco-acr-15	-1.57				

Hco-lgc-50	-1.06					*Hco-lgc-34*	-1.26			
Hco-acr-8	-1.06	*Hco-acr-8*	-1.3	*Hco-acr-8*	-1.3	*Hco-acc-2*	-1.13			
Hco-acr-7	-1.05					*Hco-acr-11*	-1.13			
Hco-acr-15	-1.02					*Hco-lgc-45*	-1.01			
Hco-lgc-42	-1.01									

ABC Transporters

Up-regulated in MHco4(WRS)

N2		N5		N5 LPGs	
Hco-abt-12	1.65				
Hco-abt-10	1.48				
Hco-mrp-7	1.25				
Hco-abt-11	1.11				
Hco-abcf-1	1.02	*Hco-abcf-1*	1.02	*Hco-abcf-1*	1.02
Hco-mrp-4	1.01				

Up-regulated in MHco10(CAVR)

N2		N5		N5 LPGs	
Hco-wht-4	2.68	*Hco-wht-4*	2.07	*Hco-wht-4*	2.07
Hco-mrp-3	2.54	*Hco-mrp-3*	1.71		
Hco-pgp-11	1.94	*Hco-pgp-11*	1.26	*Hco-pgp-11*	1.26
Hco-abt-10	1.85	*Hco-abt-10*	1.07	*Hco-abt-10*	1.07
Hco-wht-2	1.71	*Hco-wht-2*	1.31	*Hco-wht-2*	1.31
Hco-ced-7	1.5	*Hco-ced-7*	1.11	*Hco-ced-7*	1.11
Hco-wht-5.2	1.42	*Hco-wht-5.2*	1.13	*Hco-wht-5.2*	1.13
Hco-haf-9	1.38				
Hco-pgp-10	1.28				
Hco-abt-12	1.28				
Hco-mrp-4	1.27				
Hco-wht-5.1	1.14				
Hco-abt-11	1.1				

Down-regulated in MHco4(WRS)

N2		N5		N5 LPGs	
Hco-pmp-4	-1.89	*Hco-pmp-4*	-1.03		
Hco-pmp-6	-1.8	*Hco-pmp-6*	-2.06	*Hco-pmp-6*	-2.06
Hco-abt-4	-1.59	*Hco-abt-4*	-1.21		
Hco-abcf-1X	-1.53	*Hco-abcf-1X*	-1.23	*Hco-abcf-1X*	-1.23
Hco-pgp-9	-1.45	*Hco-pgp-9*	-1.13		
Hco-abt-2	-1.4				
Hco-pgp-16	-1.33	*Hco-pgp-16*	-1.38		
Hco-pgp-11	-1.32	*Hco-pgp-11*	-1.3	*Hco-pgp-11*	-1.3
Hco-pmp-2	-1.29	*Hco-pmp-2*	-1.24	*Hco-pmp-2*	-1.24
Hco-abt-7	-1.07				

Down-regulated in MHco10(CAVR)

N2		N5		N5 LPGs	
Hco-pmp-6	-2.11	*Hco-pmp-6*	-1.77	*Hco-pmp-6*	-1.77
Hco-abcf-1X	-1.9	*Hco-abcf-1X*	-1.36		
Hco-abt-2	-1.76	*Hco-abt-2*	-1.36		
Hco-pmp-4	-1.72	*Hco-pmp-4*	-1		
Hco-pgp-16	-1.57				
Hco-pgp-3	-1.5				
Hco-pgp-9	-1.24				
Hco-pmp-7	-1.02				

Supplementary Tables.

Supplementary Table S1. Data point values associated with Figure 4.

	MHco3(ISE)		MHco4(WRS)		MHco10(CAVR)	
SNP category	N2 to N5	N5 to N10	N2 to N5	N5 to N10	N2 to N5	N5 to N10
>5%	1.71	0.90	2.00	1.31	1.86	1.29
2 to 5%	1.81	0.73	1.70	1.01	1.71	0.99
1 to 2 %	1.67	0.67	1.36	0.91	1.33	0.89
.5 to 1 %	1.43	0.65	1.13	0.82	1.14	0.90
<.5%	1.30	0.63	1.02	0.81	1.02	0.82
0	1.24	0.62	0.94	0.70	0.96	0.72

Supplementary Table S2. A) Genes were classified based on their SNP rate difference in MHco4(WRS) relative to MHco3(ISE). Genes were grouped into seven SNP rate difference categories from extreme rates of 5 to 15% in both directions, to genes showing SNP rate differences of zero. The total number of genes, and the numbers within this total classified by *DESeq2* as: unexpressed, showing low counts, and the number expressed are shown for genes of each SNP rate difference category at SNP mapping allowances N2, N5, N10. Of expressed genes, the number of genes showing no differential expression (DE) in MHco4(WRS) vs. MHco3(ISE), and the number of genes differentially expressed in each of five different magnitudes - from > log2 2X fold-change up-regulated, to > log2 2X fold-change down-regulated - are shown. The mean (net) log2 fold-difference in expression (NDE), representing the average of difference in expression values of all expressed genes in each SNP rate difference category are shown at all SNP mapping allowances N2, N5, N10. B) Compiled numbers for genes of all categories are shown, contrasted by compiled numbers for genes of only the low-polymorphic gene categories, i.e. both 0 to 2% categories and the 0% category.

A.

	5 to 15% higher in MHco4(WRS)			2 to 5% higher in MHco4(WRS)			0 to 2 % higher in MHco4(WRS)			0%			0 to 2 % higher in MHco3(ISE)			2 to 5% higher in MHco3(ISE)			5 to 15% higher in MHco3(ISE)		
	N2	N5	N10	N2	N5	N10	N2	N5	N10	N2	N5	N10	N2	N5	N10	N2	N5	N10	N2	N5	N10
Total		300			3,463			12,396			3,977			4,408			407			98	
Number unexpressed	94	86	76	542	470	427	1,415	1,201	1,151	1,014	935	901	513	434	406	83	70	63	39	37	33
Number low counts	124	144	150	1,002	1,210	1,211	3,398	4,044	4,035	1,478	1,711	1,711	1,249	1,424	1,391	141	166	167	30	35	38
Number expressed	82	70	74	1,919	1,783	1,825	7,583	7,151	7,210	1,485	1,331	1,365	2,646	2,550	2,611	183	171	177	29	26	27
Number showing no DE	42	47	47	1,019	1,078	1,264	5,029	5,086	5,342	1,000	1,008	1,103	1,655	1,806	1,955	131	129	133	22	17	22
UP > log2 2X	0	1	1	6	1	1	44	27	20	10	4	3	30	11	6	4	3	2	2	0	0
UP < log2 2X, > log2 1X	1	1	1	12	26	36	229	166	157	99	45	32	199	96	85	26	14	11	2	4	2
DE under the log2 1X threshold	0	0	0	250	303	281	1,420	1,287	1,222	313	204	169	586	516	471	12	19	24	2	5	2
DOWN < log2 2X, > log2 1X	14	10	18	429	302	211	660	515	431	51	64	55	133	104	85	7	5	6	1	0	1
DOWN > log2 2X	25	11	7	203	73	32	201	70	38	12	6	3	43	17	9	3	1	1	0	0	0
Net log2 fold difference in expression (NDE)	-1.331	-1.004	-0.682	-0.846	-0.51	-0.316	-0.174	-0.13	-0.096	0.184	0.035	-0.01	0.128	0.06	0.03	0.323	0.234	0.162	0.307	0.305	0.145

B.

	Full Totals			Within 2% totals		
	N2	N5	N10	N2	N5	N10
Total		25,049			20,781	

Number unexpressed	3,700	3,233	3,057	2,942	2,570	2,458
Number low counts	7,422	8,734	8,703	6,125	7,179	7,137
Number expressed	13,927	13,082	13,289	11,714	11,032	11,186
Number no DE	8,898	9,171	9,866	7,684	7,900	8,400
UP > log2 2X	96	47	33	84	42	29
UP < log2 2X, > log2 1X	568	352	324	527	307	274
DE under the log2 1X threshold	2,583	2,334	2,169	2,319	2,007	1,862
DOWN < log2 2X, > log2 1X	1,295	1,000	807	844	683	571
DOWN > log2 2X	487	178	90	256	93	50
Net log2 fold differences in expression (NDE)	-1.409	-1.01	-0.767	0.138	-0.035	-0.076

Supplementary Table S3. Total number of differentially expressed genes (with adjusted p-values < 0.05 as determined by *DESeq2*) observed in each pairwise strain comparison, at each of the three different map allowances (N2, N5, N10). The number of genes differentially expressed at both > 1 log2 fold, and > 2 log2 fold thresholds are shown. The number of genes up- and down-regulated are also shown along with totals of both.

	MHco4(WRS) vs. MHco3(ISE)			MHco10(CAVR) vs. MHco3(ISE)			MHco4(WRS) vs. MHco10(CAVR)		
	N2	N5	N10	N2	N5	N10	N2	N5	N10
> 1 log2 fold up-reg.	664	399	355	1178	834	734	1011	447	302
> 1 log2 fold down-reg.	1783	1188	897	2282	1473	1116	968	544	442
Total > 1 log2 fold	2447	1587	1252	3460	2307	1850	1979	991	744
> 2 log2 fold up-reg.	96	46	33	288	146	97	264	59	32
> 2 log2 fold down-reg.	487	179	90	833	324	189	206	77	41
Total > 2 log2 fold	583	225	123	1121	470	286	470	136	73

Supplementary Table S4. *H. contortus* low-polymorphic gene models present in the UniProt Knowledgebase that are differentially expressed at > 4X fold-change in each pairwise strain comparison. Log2 fold differential expression is shown with length of protein sequence along side 'Protein names' and 'Gene ontology (GO)' descriptors as denoted on the UniProt Knowledgebase.

A. Up-regulated in MHco4(WRS) vs MHco3(ISE)

Gene ID	log2 fold D.E.	Length	Protein names	Gene ontology (GO)
HCOI_00655600	4.67	110	Zinc finger domain containing protein	zinc ion binding [GO:0008270]
HCOI_00444600	3.79	359	7TM GPCR domain containing protein	integral component of membrane [GO:0016021]; G-protein coupled receptor activity [GO:0004930]
HCOI_01404300	3.49	1235	Dsec\GM13241-PA	
HCOI_00355800	3.45	293	Uncharacterized protein	integral component of membrane [GO:0016021]
HCOI_00088900	3.27	361	Protein UNC-2, isoform c	calcium ion binding [GO:0005509]
HCOI_01590000	2.96	83	Uncharacterized protein	integral component of membrane [GO:0016021]
HCOI_01431800	2.8	342	Peptidase C1A domain containing protein	cysteine-type peptidase activity [GO:0008234]
HCOI_00634600	2.76	231	Uncharacterized protein (Fragment)	integral component of membrane [GO:0016021]; G-protein coupled receptor activity [GO:0004930]

HCOI_00162900	2.74	525	LGC-55 Ligand-gated chloride channel (Neurotransmitter-gated ion-channel ligand-binding and Neurotransmitter-gated ion-channel transmembrane region domain containing protein)	cell junction [GO:0030054]; integral component of membrane [GO:0016021]; plasma membrane [GO:0005886]; synapse [GO:0045202]; extracellular ligand-gated ion channel activity [GO:0005230]
HCOI_00295200	2.69	252	Glycoside hydrolase domain containing protein	integral component of membrane [GO:0016021]; lysozyme activity [GO:0003796]; carbohydrate metabolic process [GO:0005975]; cell wall macromolecule catabolic process [GO:0016998]; peptidoglycan catabolic process [GO:0009253]
HCOI_01942400	2.66	459	Carboxylic ester hydrolase (EC 3.1.1.-)	cholinesterase activity [GO:0004104]
HCOI_02043200	2.63	289	Protein OSR-1	
HCOI_02045800	2.59	209	Tau-tubulin kinase 1	ATP binding [GO:0005524]; protein kinase activity [GO:0004672]
HCOI_00007900	2.58	180	Zinc finger domain containing protein	zinc ion binding [GO:0008270]
HCOI_00063700	2.54	118	Protein F54D5.4 (Fragment)	
HCOI_01355500	2.54	711	Glutamate receptor and Ionotropic glutamate receptor domain containing protein	cell junction [GO:0030054]; integral component of membrane [GO:0016021]; postsynaptic membrane [GO:0045211]; extracellular-glutamate-gated ion channel activity [GO:0005234]; ionotropic glutamate receptor activity [GO:0004970]
HCOI_00293100	2.52	459	Armadillo/beta-catenin-like repeat family	
HCOI_01022300	2.47	87	Uncharacterized protein	
HCOI_00007400	2.43	141	Uncharacterized protein	
HCOI_00475800	2.41	1593	Uncharacterized protein	
HCOI_01789300	2.37	102	Uncharacterized protein	
HCOI_00142200	2.36	1005	Ion transport and Voltage-dependent calcium channel domain containing protein (Fragment)	integral component of membrane [GO:0016021]; voltage-gated ion channel activity [GO:0005244]
HCOI_02042400	2.31	467	Transporter	integral component of membrane [GO:0016021]; neurotransmitter:sodium symporter activity [GO:0005328]
HCOI_01832400	2.29	260	7TM GPCR domain containing protein (Fragment)	integral component of membrane [GO:0016021]; G-protein coupled receptor activity [GO:0004930]
HCOI_01737000	2.27	362	Uncharacterized protein	
HCOI_02003600	2.22	178	Protein VAP-1, isoform a	
HCOI_01416100	2.16	243	Calponin actin-binding domain containing protein	
HCOI_00793300	2.12	188	GYF domain containing protein	
HCOI_01272300	2.1	785	Zinc finger domain containing protein	zinc ion binding [GO:0008270]
HCOI_01504900	2.1	326	Uncharacterized protein (Fragment)	
HCOI_00762800	2.07	128	Ubiquitin and Ribosomal protein L40e domain containing protein (Uncharacterized protein)	ribosome [GO:0005840]; structural constituent of ribosome [GO:0003735]; translation [GO:0006412]
HCOI_02147600	2.05	243	Uncharacterized protein	
HCOI_02159500	2.05	847	Glutamine amidotransferase and Asparagine synthase domain containing protein	asparagine synthase (glutamine-hydrolyzing) activity [GO:0004066]; transferase activity [GO:0016740]; asparagine biosynthetic process [GO:0006529]; glutamine metabolic process [GO:0006541]
HCOI_02043300	2.02	315	Uncharacterized protein	
HCOI_00850700	2.02	1121	Pleckstrin homology and Unconventional myosin plant kinesin protein non-motor protein conserved region MyTH4 and FERM central domain containing protein	cytoskeleton [GO:0005856]
HCOI_01875600	2.01	972	Diaphanous GTPase-binding and Diaphanous FH3 and Actin-binding FH2 domain containing protein	actin cytoskeleton organization [GO:0030036]

B. Down-regulated in MHco4(WRS) vs MHco3(ISE)

Gene ID	log2 fold D.E.	Length	Protein names	Gene ontology (GO)

HCOI_01253600	-4.73	313	Uncharacterized protein	nucleus [GO:0005634]
HCOI_00683900	-3.57	319	Protein ZC15.5	
HCOI_01170600	-3.37	405	Beta-lactamase-related domain containing protein	
HCOI_00095600	-3.19	179	Zinc finger domain containing protein	metal ion binding [GO:0046872]; nucleic acid binding [GO:0003676]
HCOI_00719200	-3	289	Uncharacterized protein	
HCOI_02033000	-2.98	329	Peptidase A1 domain containing protein	aspartic-type endopeptidase activity [GO:0004190]
HCOI_00480800	-2.97	554	Peptidase M8 domain containing protein	membrane [GO:0016020]; metalloendopeptidase activity [GO:0004222]; cell adhesion [GO:0007155]
HCOI_00676600	-2.91	506	Bestrophin domain containing protein	
HCOI_01400400	-2.9	99	Uncharacterized protein	DNA binding [GO:0003677]
HCOI_00719400	-2.87	312	Uncharacterized protein	
HCOI_00915600	-2.8	332	Uncharacterized protein	integral component of membrane [GO:0016021]
HCOI_00035800	-2.8	142	Uncharacterized protein	
HCOI_01910900	-2.78	230	Uncharacterized protein	
HCOI_01997300	-2.75	137	Uncharacterized protein	
HCOI_00694400	-2.7	150	Uncharacterized protein	
HCOI_01240800	-2.7	136	Uncharacterized protein	
HCOI_01830200	-2.69	364	Ferric reductase transmembrane component domain containing protein (Fragment)	integral component of membrane [GO:0016021]; calcium ion binding [GO:0005509]
HCOI_00998100	-2.69	474	Major facilitator superfamily MFS-1 domain containing protein	integral component of membrane [GO:0016021]; transmembrane transport [GO:0055085]
HCOI_01771500	-2.66	101	Nematode insulin-related peptide domain containing protein	
HCOI_01584300	-2.65	484	Protein-tyrosine phosphatase domain containing protein	protein tyrosine phosphatase activity [GO:0004725]
HCOI_00905000	-2.65	186	Protein VAP-1, isoform a	
HCOI_00321000	-2.6	386	Nitrilase cyanide hydratase and apolipoprotein N-acyltransferase domain containing protein	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds [GO:0016810]; transferase activity, transferring acyl groups [GO:0016746]; nitrogen compound metabolic process [GO:0006807]
HCOI_01514300	-2.56	135	Uncharacterized protein	
HCOI_00671500	-2.56	317	Aldo keto reductase domain containing protein	oxidoreductase activity [GO:0016491]
HCOI_01408000	-2.54	235	Uncharacterized protein	integral component of membrane [GO:0016021]
HCOI_00438400	-2.54	291	C-type lectin domain containing protein	carbohydrate binding [GO:0030246]
HCOI_00886600	-2.53	318	C. briggsae CBR-OSM-7 protein	
HCOI_02082500	-2.53	289	Uncharacterized protein	mitochondrion [GO:0005739]; glycine N-acyltransferase activity [GO:0047961]
HCOI_00360200	-2.51	135	15 kDa excretory/secretory protein	
HCOI_01294600	-2.51	246	Uncharacterized protein	
HCOI_01950000	-2.5	317	Ion transport 2 domain containing protein	integral component of membrane [GO:0016021]
HCOI_00463000	-2.48	804	Uncharacterised kinase D1044.1 domain containing protein	kinase activity [GO:0016301]
HCOI_00373200	-2.46	83	Uncharacterized protein	ATP binding [GO:0005524]; DNA binding [GO:0003677]; DNA topoisomerase type II (ATP-hydrolyzing) activity [GO:0003918]; DNA topological change [GO:0006265]
HCOI_02047300	-2.45	239	PAN-1 domain containing protein (Fragment)	
HCOI_02094900	-2.44	202	SCP extracellular domain containing protein	extracellular region [GO:0005576]
HCOI_02159800	-2.42	581	Gamma-glutamyltranspeptidase domain containing protein	gamma-glutamyltransferase activity [GO:0003840]; glutathione metabolic process [GO:0006749]

HCOI_00007300	-2.37	418	Uncharacterized protein	
HCOI_01235600	-2.37	196	CT20 domain containing protein	H4/H2A histone acetyltransferase complex [GO:0043189]; regulation of transcription, DNA-templated [GO:0006355]
HCOI_02105800	-2.33	153	Uncharacterized protein	
HCOI_00915100	-2.32	147	Heat shock protein Hsp20 domain containing protein	
HCOI_00839100	-2.32	144	Activation associated secreted protein	
HCOI_00360100	-2.31	135	p15	
HCOI_00708200	-2.31	145	Uncharacterized protein	
HCOI_00209900	-2.31	139	Heat shock protein Hsp20 domain containing protein	
HCOI_00908800	-2.3	179	Nematode fatty acid retinoid binding domain containing protein	lipid binding [GO:0008289]
HCOI_01021400	-2.28	195	Uncharacterized protein	
HCOI_00892600	-2.28	318	C-type lectin and Fibrinogen domain containing protein	carbohydrate binding [GO:0030246]
HCOI_01950100	-2.28	434	CRE-TWK-11 protein	
HCOI_00998000	-2.25	164	Glutathione S-transferase domain containing protein (Fragment)	transferase activity [GO:0016740]
HCOI_01840900	-2.23	198	Uncharacterized protein	
HCOI_01414000	-2.2	133	Protein CDR-4	
HCOI_00023600	-2.19	284	Elongation of very long chain fatty acids protein (EC 2.3.1.199) (Very-long-chain 3-oxoacyl-CoA synthase)	integral component of membrane [GO:0016021]; transferase activity [GO:0016740]; fatty acid biosynthetic process [GO:0006633]
HCOI_01772400	-2.16	175	Globin domain containing protein	heme binding [GO:0020037]; iron ion binding [GO:0005506]; oxygen binding [GO:0019825]; oxygen transporter activity [GO:0005344]
HCOI_00955600	-2.15	517	Sodium:dicarboxylate symporter domain containing protein	integral component of membrane [GO:0016021]; sodium:dicarboxylate symporter activity [GO:0017153]
HCOI_01705900	-2.15	318	Peptidase S16 domain containing protein	ATP binding [GO:0005524]; ATP-dependent peptidase activity [GO:0004176]; serine-type endopeptidase activity [GO:0004252]; protein catabolic process [GO:0030163]
HCOI_00209700	-2.15	140	Heat shock protein Hsp20 domain containing protein	
HCOI_00090500	-2.14	593	Uncharacterized protein	integral component of membrane [GO:0016021]
HCOI_01522400	-2.14	296	Peptidase M12A domain containing protein	metalloendopeptidase activity [GO:0004222]
HCOI_01953600	-2.14	150	Protein F10E7.6	
HCOI_01262900	-2.12	580	BTB:POZ and BTB Kelch-associated and Kelch repeat type 1 domain containing protein	
HCOI_00031900	-2.11	204	Uncharacterized protein	
HCOI_02174400	-2.11	346	7TM GPCR domain containing protein	integral component of membrane [GO:0016021]; G-protein coupled peptide receptor activity [GO:0008528]
HCOI_00296600	-2.1	252	Uncharacterized protein	
HCOI_01772600	-2.1	175	Globin domain containing protein	heme binding [GO:0020037]; iron ion binding [GO:0005506]; oxygen binding [GO:0019825]; oxygen transporter activity [GO:0005344]
HCOI_01152300	-2.08	617	DNA-directed RNA polymerase subunit (EC 2.7.7.6) (Fragment)	DNA binding [GO:0003677]; DNA-directed RNA polymerase activity [GO:0003899]; transcription, DNA-templated [GO:0006351]
HCOI_00927000	-2.06	187	Protease inhibitor I4 domain containing protein	extracellular space [GO:0005615]; peptidase activity [GO:0008233]
HCOI_00272400	-2.06	594	Uncharacterized protein	integral component of membrane [GO:0016021]; ATP binding [GO:0005524]; ATPase activity, coupled to transmembrane movement of substances [GO:0042626]
HCOI_02055300	-2.06	407	Uncharacterized protein	
HCOI_01440300	-2.05	351	Protease inhibitor I4 domain containing protein	extracellular space [GO:0005615]; peptidase activity [GO:0008233]
HCOI_01791600	-2.05	426	Bromodomain transcription factor and Transcription factor TFIID domain containing protein	

HCOI_00724000	-2.05	134	Uncharacterized protein	integral component of membrane [GO:0016021]
HCOI_00709500	-2.04	168	Uncharacterized protein	
HCOI_00651500	-2.03	449	SCP extracellular domain containing protein	extracellular region [GO:0005576]
HCOI_01272800	-2.03	185	Apyrase domain containing protein	calcium ion binding [GO:0005509]; pyrophosphatase activity [GO:0016462]
HCOI_00437500	-2.02	160	Uncharacterized protein	
HCOI_02130500	-2.02	216	Peptidase C13 domain containing protein	GPI-anchor transamidase complex [GO:0042765]; GPI-anchor transamidase activity [GO:0003923]; peptidase activity [GO:0008233]; attachment of GPI anchor to protein [GO:0016255]
HCOI_01295900	-2.02	237	Metridin ShK toxin domain containing protein	
HCOI_01166100	-2.01	619	Semaphorin CD100 antigen and Plexin domain containing protein	integral component of membrane [GO:0016021]
HCOI_00971500	-2.01	488	Major facilitator superfamily MFS-1 domain containing protein	integral component of membrane [GO:0016021]; transmembrane transport [GO:0055085]
HCOI_00849700	-2.01	264	Uncharacterized protein	
HCOI_00716400	-2	254	Peptidase M12A domain containing protein	metalloendopeptidase activity [GO:0004222]

C. Up-regulated in MHco10(CAVR) vs MHco3(ISE)

Gene ID	log2 fold D.E.	Length	Protein names	Gene ontology (GO)
HCOI_00418900	5.87	417	Uncharacterized protein (Fragment)	integral component of membrane [GO:0016021]; extracellular ligand-gated ion channel activity [GO:0005230]
HCOI_01337900	4	174	Uncharacterised kinase D1044.1 domain containing protein	kinase activity [GO:0016301]
HCOI_02000700	3.6	260	Uncharacterized protein	
HCOI_01022300	3.47	87	Uncharacterized protein	
HCOI_00456700	3.41	112	Uncharacterized protein (Fragment)	
HCOI_02043300	3.29	315	Uncharacterized protein	
HCOI_01956700	3.25	219	7TM GPCR domain containing protein	integral component of membrane [GO:0016021]
HCOI_00153600	3.25	365	Protein FRPR-7	integral component of membrane [GO:0016021]
HCOI_02042500	3.18	181	Uncharacterized protein	
HCOI_01602500	3.14	217	Uncharacterized protein	integral component of membrane [GO:0016021]
HCOI_01167300	3.13	764	Uncharacterized protein	
HCOI_00763200	3.07	143	Uncharacterized protein	
HCOI_01737000	3.05	362	Uncharacterized protein	
HCOI_00881700	3.04	310	Protein F16C3.2	
HCOI_01401500	2.97	300	Protein K08E7.5, isoform c (Fragment)	
HCOI_00190000	2.96	69	Stem cell self-renewal protein Piwi domain containing protein	
HCOI_00444600	2.92	359	7TM GPCR domain containing protein	integral component of membrane [GO:0016021]; G-protein coupled receptor activity [GO:0004930]
HCOI_01337800	2.91	151	Uncharacterised kinase D1044.1 domain containing protein	kinase activity [GO:0016301]
HCOI_01226300	2.84	959	Uncharacterized protein	aspartic-type endopeptidase activity [GO:0004190]; nucleic acid binding [GO:0003676]; zinc ion binding [GO:0008270]
HCOI_00412100	2.83	257	Uncharacterized protein	
HCOI_00803300	2.82	238	Uncharacterized protein	integral component of membrane [GO:0016021]
HCOI_00867100	2.8	770	C. briggsae CBR-SID-1 protein	integral component of membrane [GO:0016021]; RNA transmembrane transporter activity [GO:0051033]; dsRNA transport [GO:0033227]

HCOI_00441100	2.77	1000	Lipoxygenase and Polycystin cation channel domain containing protein	integral component of membrane [GO:0016021]; calcium ion binding [GO:0005509]
HCOI_00471600	2.72	184	Protein W03D8.11	
HCOI_01543000	2.67	284	Galectin	carbohydrate binding [GO:0030246]
HCOI_00926900	2.62	291	Protease inhibitor I4 domain containing protein	extracellular space [GO:0005615]; peptidase activity [GO:0008233]
HCOI_00513400	2.62	536	Protein T08B6.4	integral component of membrane [GO:0016021]
HCOI_00803400	2.62	201	Uncharacterized protein	integral component of membrane [GO:0016021]
HCOI_02172000	2.56	202	Lipase domain containing protein	hydrolase activity [GO:0016787]
HCOI_00771000	2.55	813	Dvir\GJ11255-PA	
HCOI_00057100	2.54	773	EGF receptor domain containing protein	integral component of membrane [GO:0016021]
HCOI_01696500	2.49	447	RNA-directed DNA polymerase (Reverse transcriptase) domain containing protein	RNA-directed DNA polymerase activity [GO:0003964]
HCOI_01322900	2.46	80	Uncharacterized protein	
HCOI_00262600	2.45	166	Microsomal aminopeptidase (Microsomal aminopeptidase H11) (Fragment)	aminopeptidase activity [GO:0004177]
HCOI_01899200	2.43	625	Serine/threonine-protein phosphatase (EC 3.1.3.16)	calcium ion binding [GO:0005509]; iron ion binding [GO:0005506]; manganese ion binding [GO:0030145]; phosphoprotein phosphatase activity [GO:0004721]; detection of stimulus involved in sensory perception [GO:0050906]
HCOI_00665300	2.43	171	Uncharacterized protein	
HCOI_01706200	2.42	724	CBN-PQN-46 protein	
HCOI_01106200	2.41	833	Serine-rich adhesin for platelets family	
HCOI_02043200	2.4	289	Protein OSR-1	
HCOI_00919200	2.39	349	LBP/BPI/CETP family domain-containing protein	lipid binding [GO:0008289]
HCOI_02051200	2.39	194	Protein ZK675.4	
HCOI_01840000	2.39	104	Uncharacterized protein	
HCOI_00040300	2.36	371	Protein CYN-17, isoform a	
HCOI_00919100	2.35	196	Protein C06G1.1, isoform a	lipid binding [GO:0008289]
HCOI_02098400	2.34	280	Uncharacterized protein	
HCOI_00441800	2.34	112	Uncharacterized protein	
HCOI_01776700	2.34	210	Uncharacterized protein	integral component of membrane [GO:0016021]
HCOI_00162900	2.33	525	LGC-55, Ligand-gated chloride channel (Neurotransmitter-gated ion-channel ligand-binding and Neurotransmitter-gated ion-channel transmembrane region domain containing protein)	cell junction [GO:0030054]; integral component of membrane [GO:0016021]; plasma membrane [GO:0005886]; synapse [GO:0045202]; extracellular ligand-gated ion channel activity [GO:0005230]
HCOI_00180600	2.33	148	Uncharacterized protein	integral component of membrane [GO:0016021]
HCOI_00137800	2.31	71	Uncharacterized protein	
HCOI_01958000	2.28	386	Uncharacterized protein	
HCOI_02039400	2.26	1527	Immunoglobulin I-set domain containing protein (Fragment)	
HCOI_00260400	2.26	223	Protein T05A7.1	
HCOI_00394900	2.25	841	Selectin-like protein (Fragment)	carbohydrate binding [GO:0030246]
HCOI_01492900	2.24	427	BTB:POZ and BTB Kelch-associated and Kelch repeat type 1 domain containing protein	
HCOI_02042400	2.24	467	Transporter	integral component of membrane [GO:0016021]; neurotransmitter:sodium symporter activity [GO:0005328]
HCOI_00930900	2.23	261	ELL-associated factor domain containing protein	ELL-EAF complex [GO:0032783]; regulation of transcription, DNA-templated [GO:0006355]

HCOI_02167300	2.23	124	Uncharacterized protein	
HCOI_01901700	2.2	401	Transcription factor jumonji domain containing protein	
HCOI_01226200	2.19	583	Integrase domain containing protein	nucleic acid binding [GO:0003676]; DNA integration [GO:0015074]
HCOI_01140400	2.18	67	Uncharacterized protein	
HCOI_00293100	2.18	459	Armadillo/beta-catenin-like repeat family	
HCOI_00582100	2.17	476	Uncharacterized protein	
HCOI_00980400	2.17	326	Uncharacterized protein	
HCOI_01825500	2.16	193	Nucleolar protein 10 (Fragment)	
HCOI_02042900	2.16	565	Transporter	integral component of membrane [GO:0016021]; neurotransmitter:sodium symporter activity [GO:0005328]
HCOI_01438100	2.15	430	Uncharacterized protein	integral component of membrane [GO:0016021]
HCOI_01428100	2.15	277	Uncharacterized protein	
HCOI_01590000	2.14	83	Uncharacterized protein	integral component of membrane [GO:0016021]
HCOI_00881600	2.14	358	Nematode cuticle collagen and Collagen triple helix repeat domain containing protein	collagen trimer [GO:0005581]; integral component of membrane [GO:0016021]; structural constituent of cuticle [GO:0042302]
HCOI_01621500	2.13	380	Amino acid transporter domain containing protein (Fragment)	integral component of membrane [GO:0016021]
HCOI_00260300	2.12	120	Parasitic stage specific protein 1	
HCOI_01516900	2.12	216	Uncharacterized protein (Fragment)	mitochondrion [GO:0005739]; double-stranded DNA binding [GO:0003690]; regulation of transcription, DNA-templated [GO:0006355]
HCOI_00696400	2.12	119	DNA-directed RNA polymerase subunit	nucleolus [GO:0005730]; DNA-directed RNA polymerase activity [GO:0003899]; nucleic acid binding [GO:0003676]; zinc ion binding [GO:0008270]; transcription, DNA-templated [GO:0006351]
HCOI_00295200	2.12	252	Glycoside hydrolase domain containing protein	integral component of membrane [GO:0016021]; lysozyme activity [GO:0003796]; carbohydrate metabolic process [GO:0005975]; cell wall macromolecule catabolic process [GO:0016998]; peptidoglycan catabolic process [GO:0009253]
HCOI_00088800	2.11	683	Ion transport domain containing protein	voltage-gated calcium channel complex [GO:0005891]; voltage-gated calcium channel activity [GO:0005245]
HCOI_00408700	2.11	393	Ion transport 2 domain containing protein	integral component of membrane [GO:0016021]; potassium channel activity [GO:0005267]
HCOI_00769700	2.1	682	Solute carrier organic anion transporter family member	integral component of membrane [GO:0016021]; plasma membrane [GO:0005886]; transporter activity [GO:0005215]; ion transport [GO:0006811]
HCOI_00385800	2.1	640	Uncharacterized protein	integral component of membrane [GO:0016021]
HCOI_01545500	2.09	266	Uncharacterized protein	
HCOI_00767800	2.09	437	Uncharacterized protein	
HCOI_01169000	2.07	498	Uncharacterized protein (Fragment)	membrane [GO:0016020]; ATP binding [GO:0005524]; ATPase activity [GO:0016887]
HCOI_01255000	2.06	368	UDP-glucuronosyltransferase (EC 2.4.1.17)	integral component of membrane [GO:0016021]; glucuronosyltransferase activity [GO:0015020]; metabolic process [GO:0008152]
HCOI_00446300	2.06	495	Kinesin-like protein	microtubule [GO:0005874]; ATP binding [GO:0005524]; microtubule motor activity [GO:0003777]; microtubule-based movement [GO:0007018]
HCOI_01162700	2.05	152	Galectin (Fragment)	carbohydrate binding [GO:0030246]
HCOI_01411000	2.05	94	Uncharacterized protein	
HCOI_00655600	2.05	110	Zinc finger domain containing protein	zinc ion binding [GO:0008270]

HCOI_00792100	2.05	485	7TM GPCR domain containing protein (Fragment)	integral component of membrane [GO:0016021]; G-protein coupled receptor activity [GO:0004930]
HCOI_01429500	2.04	466	Protein R09E10.6	
HCOI_00606600	2.03	202	Uncharacterized protein	
HCOI_01881900	2.03	608	TWiK family of potassium channels protein 7	integral component of membrane [GO:0016021]; potassium channel activity [GO:0005267]
HCOI_01642000	2.03	729	Protein M28.8	integral component of membrane [GO:0016021]
HCOI_01932100	2.02	150	Uncharacterized protein	
HCOI_01984100	2.01	324	7TM GPCR domain containing protein	integral component of membrane [GO:0016021]; sensory perception of chemical stimulus [GO:0007606]

D. Down-regulated in MHco10(CAVR) vs MHco3(ISE)

Gene ID	log2 fold D.E.	Length	Protein names	Gene ontology (GO)
HCOI_01253600	-4.74	313	Uncharacterized protein	nucleus [GO:0005634]
HCOI_00719200	-3.8	289	Uncharacterized protein	
HCOI_00593400	-3.47	96	Uncharacterized protein	
HCOI_00007300	-3.36	418	Uncharacterized protein	
HCOI_01204900	-3.31	112	Ribosomal protein S32 domain containing protein	ribosome [GO:0005840]
HCOI_00221000	-3.31	113	Uncharacterized protein	
HCOI_01540300	-3.29	207	Uncharacterized protein	
HCOI_01514600	-3.17	147	Uncharacterized protein	
HCOI_01996700	-3.13	137	Uncharacterized protein	
HCOI_01997300	-3.1	137	Uncharacterized protein	
HCOI_01294600	-3.02	246	Uncharacterized protein	
HCOI_00893400	-2.97	463	Zinc finger domain containing protein	metal ion binding [GO:0046872]
HCOI_00621300	-2.94	352	Protease inhibitor I4 domain containing protein	extracellular space [GO:0005615]; peptidase activity [GO:0008233]
HCOI_00905000	-2.92	186	Protein VAP-1, isoform a	
HCOI_00938300	-2.91	386	Uncharacterized protein (Fragment)	integral component of membrane [GO:0016021]; extracellular ligand-gated ion channel activity [GO:0005230]
HCOI_00463000	-2.89	804	Uncharacterised kinase D1044.1 domain containing protein	kinase activity [GO:0016301]
HCOI_00360200	-2.89	135	15 kDa excretory/secretory protein	
HCOI_01137800	-2.89	376	Actin actin domain containing protein	ATP binding [GO:0005524]
HCOI_00694600	-2.86	135	EF hand domain containing protein	calcium ion binding [GO:0005509]
HCOI_00684000	-2.85	250	Glycoprotein-N-acetylgalactosamine	membrane [GO:0016020]; galactosyltransferase activity [GO:0008378]; protein glycosylation [GO:0006486]
HCOI_01170600	-2.82	405	Beta-lactamase-related domain containing protein	
HCOI_01497600	-2.74	230	Uncharacterized protein (Fragment)	
HCOI_01615000	-2.71	137	p15	
HCOI_01240800	-2.71	136	Uncharacterized protein	
HCOI_01388800	-2.7	84	Protein TAG-307	
HCOI_01021400	-2.69	195	Uncharacterized protein	
HCOI_00589900	-2.68	294	UDP-glucuronosyl UDP-glucosyltransferase domain containing protein (Fragment)	integral component of membrane [GO:0016021]; transferase activity, transferring hexosyl groups [GO:0016758]; metabolic process [GO:0008152]
HCOI_00271700	-2.66	200	Uncharacterized protein	
HCOI_00450400	-2.64	475	Uncharacterized protein	

HCOI_00296600	-2.64	252	Uncharacterized protein	
HCOI_00171700	-2.61	215	SCP extracellular domain containing protein	extracellular region [GO:0005576]
HCOI_00023600	-2.61	284	Elongation of very long chain fatty acids protein (EC 2.3.1.199) (Very-long-chain 3-oxoacyl-CoA synthase)	integral component of membrane [GO:0016021]; transferase activity [GO:0016740]; fatty acid biosynthetic process [GO:0006633]
HCOI_02033000	-2.6	329	Peptidase A1 domain containing protein	aspartic-type endopeptidase activity [GO:0004190]
HCOI_01584300	-2.59	484	Protein-tyrosine phosphatase domain containing protein	protein tyrosine phosphatase activity [GO:0004725]
HCOI_00524200	-2.58	346	Peptidase C1A domain containing protein	cysteine-type peptidase activity [GO:0008234]
HCOI_01248100	-2.58	272	Collagen triple helix repeat domain containing protein	collagen trimer [GO:0005581]
HCOI_00437500	-2.55	160	Uncharacterized protein	
HCOI_00621200	-2.55	132	Uncharacterized protein	
HCOI_02033100	-2.53	394	Peptidase A1 domain containing protein	aspartic-type endopeptidase activity [GO:0004190]
HCOI_02130500	-2.52	216	Peptidase C13 domain containing protein	GPI-anchor transamidase complex [GO:0042765]; GPI-anchor transamidase activity [GO:0003923]; peptidase activity [GO:0008233]; attachment of GPI anchor to protein [GO:0016255]
HCOI_00214700	-2.51	93	Endoglin CD105 antigen domain containing protein	
HCOI_01771500	-2.5	101	Nematode insulin-related peptide domain containing protein	
HCOI_02060300	-2.5	575	BRICHOS domain containing protein	integral component of membrane [GO:0016021]
HCOI_00620700	-2.49	188	Glutathione peroxidase	glutathione peroxidase activity [GO:0004602]; response to oxidative stress [GO:0006979]
HCOI_02107900	-2.48	146	Uncharacterized protein	
HCOI_00915200	-2.47	165	Heat shock protein Hsp20 domain containing protein	
HCOI_01868700	-2.45	254	Tyrosinase domain containing protein	metal ion binding [GO:0046872]; oxidoreductase activity [GO:0016491]
HCOI_00090500	-2.45	593	Uncharacterized protein	integral component of membrane [GO:0016021]
HCOI_01772500	-2.44	175	Globin domain containing protein	heme binding [GO:0020037]; iron ion binding [GO:0005506]; oxygen binding [GO:0019825]; oxygen transporter activity [GO:0005344]
HCOI_00927000	-2.43	187	Protease inhibitor I4 domain containing protein	extracellular space [GO:0005615]; peptidase activity [GO:0008233]
HCOI_01306000	-2.4	271	Aquaporin-10	integral component of membrane [GO:0016021]
HCOI_01667300	-2.39	226	Venom allergen/ancylostoma secreted protein-like	
HCOI_00919400	-2.36	183	SH2 motif domain containing protein	
HCOI_01798700	-2.35	87	RNA recognition motif domain containing protein	nucleic acid binding [GO:0003676]; nucleotide binding [GO:0000166]
HCOI_00031900	-2.35	204	Uncharacterized protein	
HCOI_00514900	-2.35	137	15 kDa excretory/secretory protein	
HCOI_01772600	-2.33	175	Globin domain containing protein	heme binding [GO:0020037]; iron ion binding [GO:0005506]; oxygen binding [GO:0019825]; oxygen transporter activity [GO:0005344]
HCOI_00839100	-2.33	144	Activation associated secreted protein	
HCOI_01262900	-2.33	580	BTB:POZ and BTB Kelch-associated and Kelch repeat type 1 domain containing protein	
HCOI_00359200	-2.29	76	Uncharacterized protein	
HCOI_00915300	-2.29	136	Heat shock protein Hsp20 domain containing protein	
HCOI_01387700	-2.29	342	Uncharacterized protein	integral component of membrane [GO:0016021]
HCOI_01202200	-2.27	502	Protein Y54G11A.1	
HCOI_00223400	-2.23	1169	Uncharacterized protein (Fragment)	
HCOI_00696100	-2.22	128	RNA polymerase Rpb6 domain containing protein	DNA-directed RNA polymerase II, core complex [GO:0005665]; DNA binding [GO:0003677]; DNA-directed RNA polymerase activity [GO:0003899]; transcription, DNA-templated [GO:0006351]
HCOI_01953600	-2.22	150	Protein F10E7.6	

HCOI_01996600	-2.21	137	15 kDa excretory/secretory protein	
HCOI_02003500	-2.21	120	Uncharacterized protein	
HCOI_00862400	-2.19	112	Metalloprotease I	metallopeptidase activity [GO:0008237]
HCOI_01449200	-2.19	449	SCP extracellular domain containing protein	extracellular region [GO:0005576]
HCOI_01615200	-2.19	139	Transthyretin domain containing protein	extracellular space [GO:0005615]
HCOI_01456200	-2.19	134	Uncharacterized protein (Fragment)	
HCOI_00388200	-2.18	420	SCP extracellular domain containing protein	extracellular region [GO:0005576]
HCOI_00330700	-2.18	567	Myosin head and IQ calmodulin-binding region and Myosin tail domain containing protein (Fragment)	myosin complex [GO:0016459]; ATP binding [GO:0005524]; motor activity [GO:0003774]
HCOI_00651500	-2.17	449	SCP extracellular domain containing protein	extracellular region [GO:0005576]
HCOI_00480800	-2.17	554	Peptidase M8 domain containing protein	membrane [GO:0016020]; metalloendopeptidase activity [GO:0004222]; cell adhesion [GO:0007155]
HCOI_00338800	-2.15	238	RNA-directed DNA polymerase (Reverse transcriptase) domain containing protein	RNA-directed DNA polymerase activity [GO:0003964]
HCOI_00927100	-2.13	698	Protein C35D10.8 (Fragment)	integral component of membrane [GO:0016021]; protein phosphatase inhibitor activity [GO:0004864]; regulation of phosphoprotein phosphatase activity [GO:0043666]; regulation of signal transduction [GO:0009966]
HCOI_00700300	-2.13	344	Proteasome component region PCI domain containing protein (Fragment)	proteasome complex [GO:0000502]
HCOI_00229700	-2.12	288	Uncharacterized protein	integral component of membrane [GO:0016021]
HCOI_00722000	-2.12	270	Dere\GG20951-PA	integral component of membrane [GO:0016021]; neurotransmitter:sodium symporter activity [GO:0005328]
HCOI_01669100	-2.12	102	Ubiquitin-related modifier 1 homolog	cytosol [GO:0005829]; protein urmylation [GO:0032447]; tRNA thio-modification [GO:0034227]; tRNA wobble uridine modification [GO:0002098]
HCOI_01152400	-2.12	139	Protein Y105C5B.5	
HCOI_01985600	-2.11	144	Uncharacterized protein	
HCOI_01190400	-2.11	218	SCP extracellular domain containing protein	extracellular region [GO:0005576]
HCOI_00791700	-2.09	220	SCP extracellular domain containing protein	extracellular region [GO:0005576]
HCOI_01659400	-2.09	112	Metalloprotease	metallopeptidase activity [GO:0008237]
HCOI_00719300	-2.08	318	Uncharacterized protein	
HCOI_00823700	-2.08	571	Metalloendopeptidase (EC 3.4.24.-)	metalloendopeptidase activity [GO:0004222]; zinc ion binding [GO:0008270]; molting cycle, collagen and cuticulin-based cuticle [GO:0018996]
HCOI_00697000	-2.07	427	FAD-dependent pyridine nucleotide-disulphide oxidoreductase domain containing protein	oxidoreductase activity [GO:0016491]
HCOI_01300800	-2.07	290	Short-chain dehydrogenase reductase SDR domain containing protein	
HCOI_00385500	-2.06	298	Nematode cuticle collagen and Collagen triple helix repeat domain containing protein	collagen trimer [GO:0005581]; integral component of membrane [GO:0016021]; structural constituent of cuticle [GO:0042302]
HCOI_01546000	-2.06	95	Uncharacterized protein	
HCOI_01272200	-2.05	1471	Uncharacterized protein (Fragment)	integral component of membrane [GO:0016021]
HCOI_02009200	-2.05	136	Uncharacterized protein	
HCOI_00709500	-2.05	168	Uncharacterized protein	
HCOI_02183800	-2.04	219	Venom allergen/ancylostoma secreted protein-like	
HCOI_01205600	-2.03	188	Uncharacterized protein	
HCOI_00462900	-2.03	409	Uncharacterised kinase D1044.1 domain containing protein	integral component of membrane [GO:0016021]; kinase activity [GO:0016301]

HCOI_01772400	-2.03	175	Globin domain containing protein	heme binding [GO:0020037]; iron ion binding [GO:0005506]; oxygen binding [GO:0019825]; oxygen transporter activity [GO:0005344]
HCOI_00955600	-2.03	517	Sodium:dicarboxylate symporter domain containing protein	integral component of membrane [GO:0016021]; sodium:dicarboxylate symporter activity [GO:0017153]
HCOI_01994500	-2.02	204	Uncharacterized protein	integral component of membrane [GO:0016021]
HCOI_01984700	-2.02	89	Uncharacterized protein	
HCOI_01772000	-2.02	297	Globin domain containing protein	integral component of membrane [GO:0016021]; heme binding [GO:0020037]; oxygen binding [GO:0019825]; oxygen transporter activity [GO:0005344]
HCOI_01986300	-2.01	215	SCP extracellular domain containing protein	extracellular region [GO:0005576]
HCOI_01451800	-2.01	107	Collagen triple helix repeat domain containing protein (Fragment)	collagen trimer [GO:0005581]
HCOI_01600100	-2	188	SCP extracellular domain containing protein	

E. Up-regulated in MHco4(WRS) vs MHco10(CAVR)

Gene ID	log2 fold D.E.	Length	Protein names	Gene ontology (GO)
HCOI_02045800	3.48	209	Tau-tubulin kinase 1	ATP binding [GO:0005524]; protein kinase activity [GO:0004672]
HCOI_00385500	3.03	298	Nematode cuticle collagen and Collagen triple helix repeat domain containing protein	collagen trimer [GO:0005581]; integral component of membrane [GO:0016021]; structural constituent of cuticle [GO:0042302]
HCOI_01236700	2.87	185	Zinc finger domain containing protein	
HCOI_01204900	2.84	112	Ribosomal protein S32 domain containing protein	ribosome [GO:0005840]
HCOI_00374400	2.79	174	ATP synthase assembly factor FMC1 domain containing protein	
HCOI_02011700	2.75	137	Uncharacterized protein	
HCOI_00454600	2.74	102	Uncharacterized protein	
HCOI_00355800	2.71	293	Uncharacterized protein	integral component of membrane [GO:0016021]
HCOI_00655600	2.62	110	Zinc finger domain containing protein	zinc ion binding [GO:0008270]
HCOI_01248100	2.58	272	Collagen triple helix repeat domain containing protein	collagen trimer [GO:0005581]
HCOI_01952000	2.56	106	RE55111p	
HCOI_01155600	2.53	83	Uncharacterized protein	ATP binding [GO:0005524]; DNA binding [GO:0003677]; DNA topoisomerase type II (ATP-hydrolyzing) activity [GO:0003918]; DNA topological change [GO:0006265]
HCOI_00621300	2.52	352	Protease inhibitor I4 domain containing protein	extracellular space [GO:0005615]; peptidase activity [GO:0008233]
HCOI_00682700	2.45	488	DNA primase domain containing protein	DNA primase activity [GO:0003896]
HCOI_01710300	2.44	1981	Dynein heavy chain and ATPase associated with various cellular activities domain containing protein	ATP binding [GO:0005524]; ATPase activity [GO:0016887]; microtubule motor activity [GO:0003777]; microtubule-based movement [GO:0007018]
HCOI_02014700	2.4	212	DNA RNA helicase domain containing protein (Fragment)	integral component of membrane [GO:0016021]; helicase activity [GO:0004386]
HCOI_00892400	2.36	371	C-type lectin and Fibrinogen domain containing protein	carbohydrate binding [GO:0030246]
HCOI_00214700	2.25	93	Endoglin CD105 antigen domain containing protein	
HCOI_01629200	2.23	1377	Uncharacterized protein	
HCOI_01788300	2.22	942	Serine threonine protein kinase-related domain containing protein	ATP binding [GO:0005524]; protein kinase activity [GO:0004672]
HCOI_02076500	2.21	108	DNA-directed RNA polymerase domain containing protein	DNA binding [GO:0003677]; DNA-directed RNA polymerase activity [GO:0003899]; transcription, DNA-templated [GO:0006351]
HCOI_02040900	2.19	118	Uncharacterized protein	
HCOI_00938300	2.17	386	Uncharacterized protein (Fragment)	integral component of membrane [GO:0016021]; extracellular ligand-gated ion channel activity [GO:0005230]

HCOI_01095600	2.14	246	Uncharacterised protein family UPF0005 domain containing protein	integral component of membrane [GO:0016021]
HCOI_00630400	2.14	121	Transcription initiation factor IID domain containing protein	translation initiation factor activity [GO:0003743]; transcription from RNA polymerase II promoter [GO:0006366]
HCOI_02003600	2.13	178	Protein VAP-1, isoform a	
HCOI_00663500	2.11	347	CoA-transferase family III domain containing protein	transferase activity [GO:0016740]
HCOI_00779000	2.08	183	Histone H1 H5 domain containing protein	nucleosome [GO:0000786]; nucleus [GO:0005634]; DNA binding [GO:0003677]; nucleosome assembly [GO:0006334]
HCOI_00007400	2.07	141	Uncharacterized protein	
HCOI_01469100	2.03	128	Uncharacterized protein	
HCOI_00087600	2.02	242	Snf7 domain containing protein	intracellular [GO:0005622]; vacuolar transport [GO:0007034]
HCOI_00259000	2.01	217	Ribosomal protein L6E domain containing protein	ribosome [GO:0005840]; structural constituent of ribosome [GO:0003735]; translation [GO:0006412]
HCOI_01330000	2.01	249	Phosphomannomutase (EC 5.4.2.8)	cytoplasm [GO:0005737]; phosphomannomutase activity [GO:0004615]; GDP-mannose biosynthetic process [GO:0009298]

F. Down-regulated in MHco4(WRS) vs MHco10(CAVR)

Gene ID	log2 fold D.E.	Length	Protein names	Gene ontology (GO)
HCOI_00418900	-5.39	417	Uncharacterized protein (Fragment)	integral component of membrane [GO:0016021]; extracellular ligand-gated ion channel activity [GO:0005230]
HCOI_01255000	-4.38	368	UDP-glucuronosyltransferase (EC 2.4.1.17)	integral component of membrane [GO:0016021]; glucuronosyltransferase activity [GO:0015020]; metabolic process [GO:0008152]
HCOI_00438300	-3.49	842	Mammalian uncoordinated homology 13 and C2 calcium-dependent membrane targeting domain containing protein	
HCOI_00400400	-3.45	131	CBN-SPP-18 protein	
HCOI_00446300	-3.3	495	Kinesin-like protein	microtubule [GO:0005874]; ATP binding [GO:0005524]; microtubule motor activity [GO:0003777]; microtubule-based movement [GO:0007018]
HCOI_01492900	-3.06	427	BTB:POZ and BTB Kelch-associated and Kelch repeat type 1 domain containing protein	
HCOI_00673300	-3.03	134	Uncharacterized protein	
HCOI_00376200	-2.86	656	DNA-binding RFX domain containing protein	DNA binding [GO:0003677]; regulation of transcription, DNA-templated [GO:0006355]
HCOI_00441100	-2.75	1000	Lipoxygenase and Polycystin cation channel domain containing protein	integral component of membrane [GO:0016021]; calcium ion binding [GO:0005509]
HCOI_00471600	-2.74	184	Protein W03D8.11	
HCOI_00260400	-2.65	223	Protein T05A7.1	
HCOI_00190000	-2.64	69	Stem cell self-renewal protein Piwi domain containing protein	
HCOI_00676600	-2.62	506	Bestrophin domain containing protein	
HCOI_00502000	-2.55	90	Uncharacterized protein (Fragment)	
HCOI_01378500	-2.54	503	C. briggsae CBR-VAB-8 protein	
HCOI_01602100	-2.54	282	7TM GPCR and RNA-directed DNA polymerase (Reverse transcriptase) domain containing protein	integral component of membrane [GO:0016021]; RNA-directed DNA polymerase activity [GO:0003964]
HCOI_00153600	-2.53	365	Protein FRPR-7	integral component of membrane [GO:0016021]
HCOI_00456700	-2.52	112	Uncharacterized protein (Fragment)	
HCOI_00513400	-2.5	536	Protein T08B6.4	integral component of membrane [GO:0016021]

HCOI_01337900	-2.5	174	Uncharacterised kinase D1044.1 domain containing protein	kinase activity [GO:0016301]
HCOI_01401500	-2.44	300	Protein K08E7.5, isoform c (Fragment)	
HCOI_01442300	-2.42	564	Uncharacterized protein	
HCOI_00608600	-2.41	646	Uncharacterized protein	
HCOI_01543000	-2.28	284	Galectin	carbohydrate binding [GO:0030246]
HCOI_01438100	-2.25	430	Uncharacterized protein	integral component of membrane [GO:0016021]
HCOI_01191500	-2.23	274	SCP extracellular domain containing protein	
HCOI_00441900	-2.22	101	Uncharacterized protein	
HCOI_01950100	-2.2	434	CRE-TWK-11 protein	
HCOI_01753400	-2.19	279	S1 domain containing protein	translation initiation factor activity [GO:0003743]
HCOI_00622400	-2.18	1250	Uncharacterized protein (Fragment)	integral component of membrane [GO:0016021]; ATP binding [GO:0005524]; ATPase activity, coupled to transmembrane movement of substances [GO:0042626]
HCOI_01052400	-2.16	224	Protease inhibitor I8 domain containing protein	peptidase activity [GO:0008233]
HCOI_02167300	-2.15	124	Uncharacterized protein	
HCOI_01932100	-2.14	150	Uncharacterized protein	
HCOI_01845500	-2.14	91	Uncharacterized protein	
HCOI_00354000	-2.12	99	Uncharacterized protein	
HCOI_01956700	-2.12	219	7TM GPCR domain containing protein	integral component of membrane [GO:0016021]
HCOI_01602300	-2.11	380	C2 calcium-dependent membrane targeting domain containing protein	membrane [GO:0016020]; synaptic vesicle [GO:0008021]; exocytosis [GO:0006887]
HCOI_00441800	-2.1	112	Uncharacterized protein	
HCOI_02172000	-2.09	202	Lipase domain containing protein	hydrolase activity [GO:0016787]
HCOI_01602200	-2.07	335	7TM GPCR domain containing protein	integral component of membrane [GO:0016021]
HCOI_01899200	-2.06	625	Serine/threonine-protein phosphatase (EC 3.1.3.16)	calcium ion binding [GO:0005509]; iron ion binding [GO:0005506]; manganese ion binding [GO:0030145]; phosphoprotein phosphatase activity [GO:0004721]; detection of stimulus involved in sensory perception [GO:0050906]
HCOI_01760800	-2.06	332	Uncharacterized protein	integral component of membrane [GO:0016021]; mitochondrion [GO:0005739]; mitochondrial electron transport, NADH to ubiquinone [GO:0006120]
HCOI_00366600	-2.06	137	Uncharacterized protein	integral component of membrane [GO:0016021]
HCOI_01597200	-2.06	431	Clc domain containing protein	integral component of membrane [GO:0016021]
HCOI_01602500	-2.05	217	Uncharacterized protein	integral component of membrane [GO:0016021]
HCOI_00084900	-2.05	319	Serine threonine protein kinase-related domain containing protein	ATP binding [GO:0005524]; protein serine/threonine kinase activity [GO:0004674]
HCOI_00040300	-2.03	371	Protein CYN-17, isoform a	
HCOI_01256400	-2.03	249	Zinc finger protein	
HCOI_00792300	-2.03	247	Protein F01F1.3 (Uncharacterized protein)	
HCOI_01931900	-2.02	501	Tyrosine-protein kinase (EC 2.7.10.2)	ATP binding [GO:0005524]; non-membrane spanning protein tyrosine kinase activity [GO:0004715]
HCOI_01904900	-2.01	179	Galectin	carbohydrate binding [GO:0030246]
HCOI_01527500	-2	364	Sterile alpha motif SAM and Sterile alpha motif homology 2 and Toll-Interleukin receptor domain containing protein	signal transduction [GO:0007165]

Figure 1 of 5 pdf, (ppt available)

Figure 1.

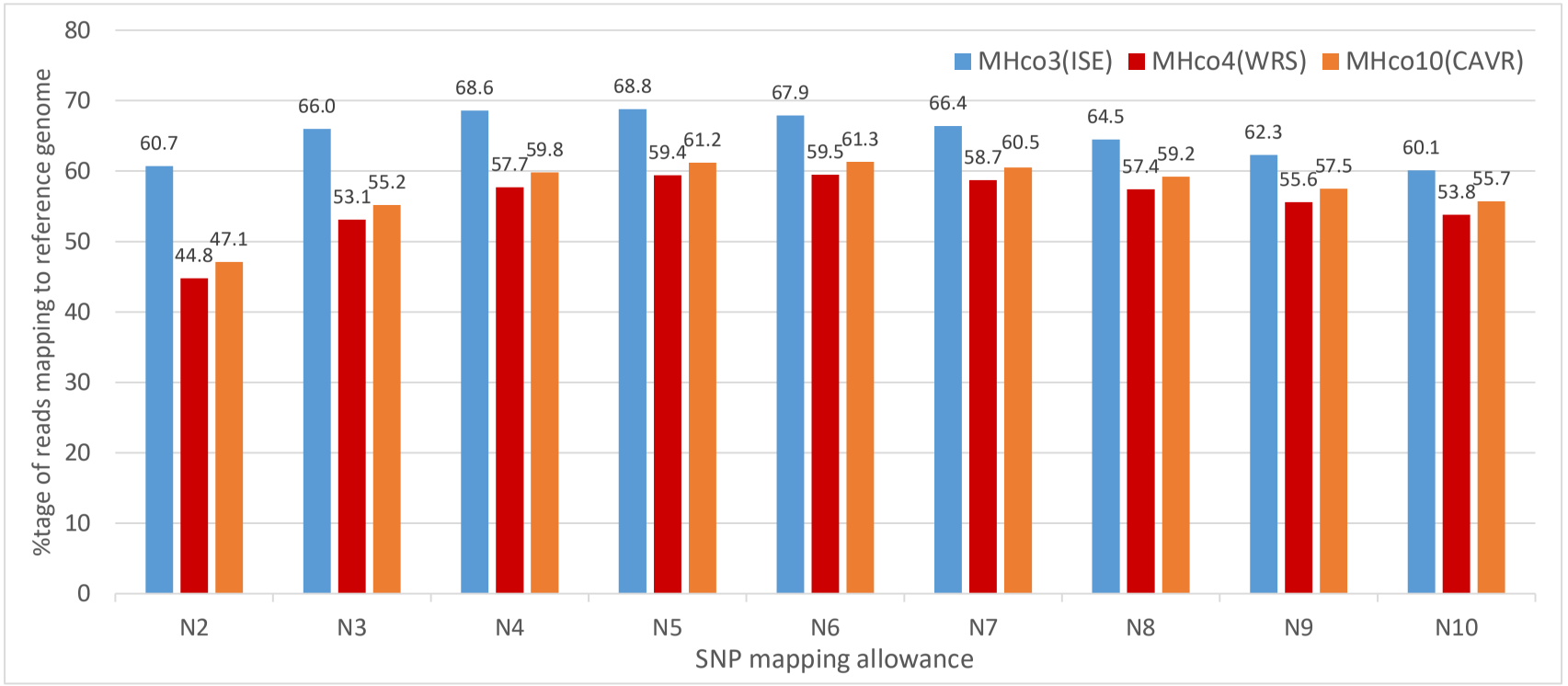


Figure 2

Figure 2

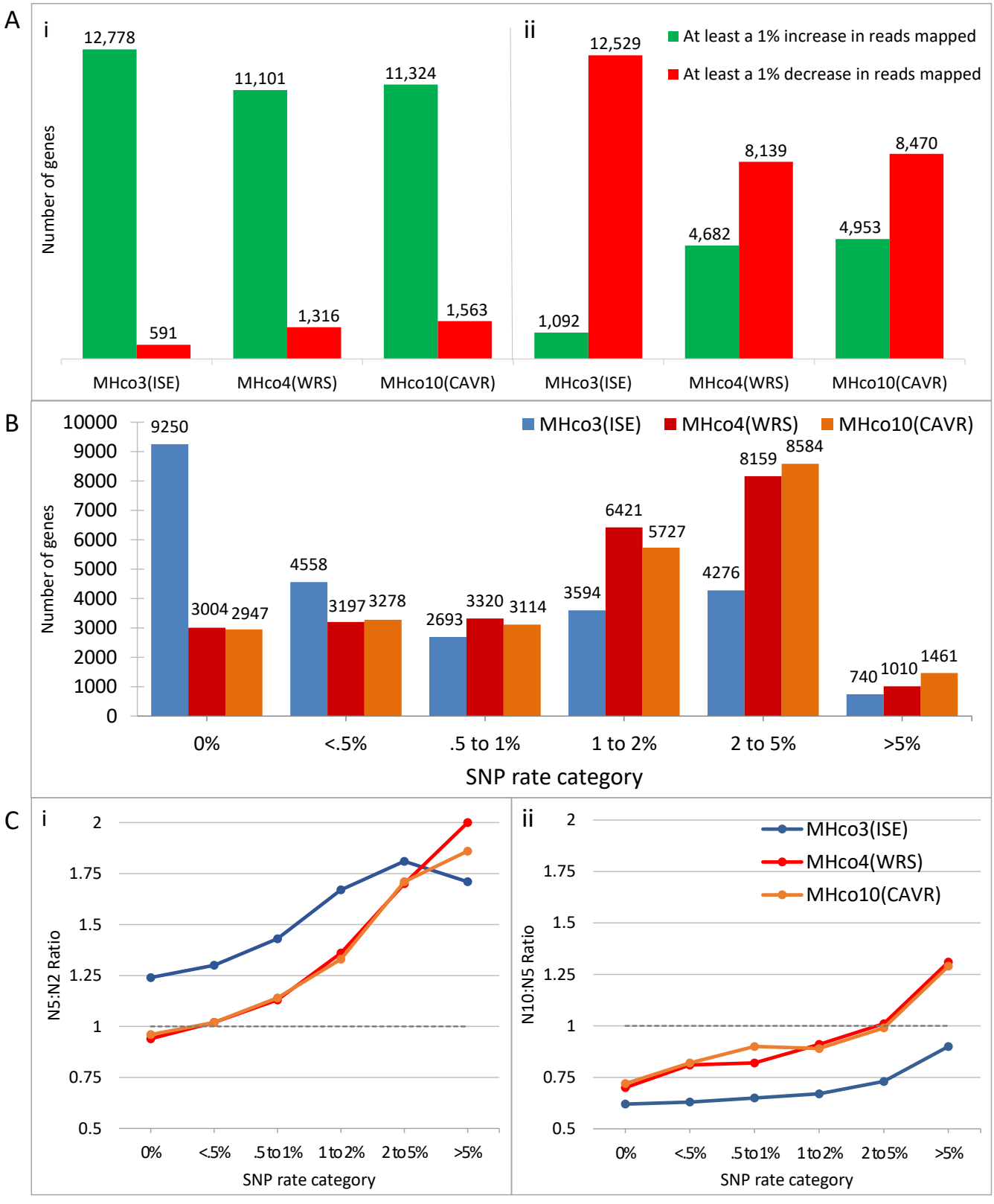


Figure 3

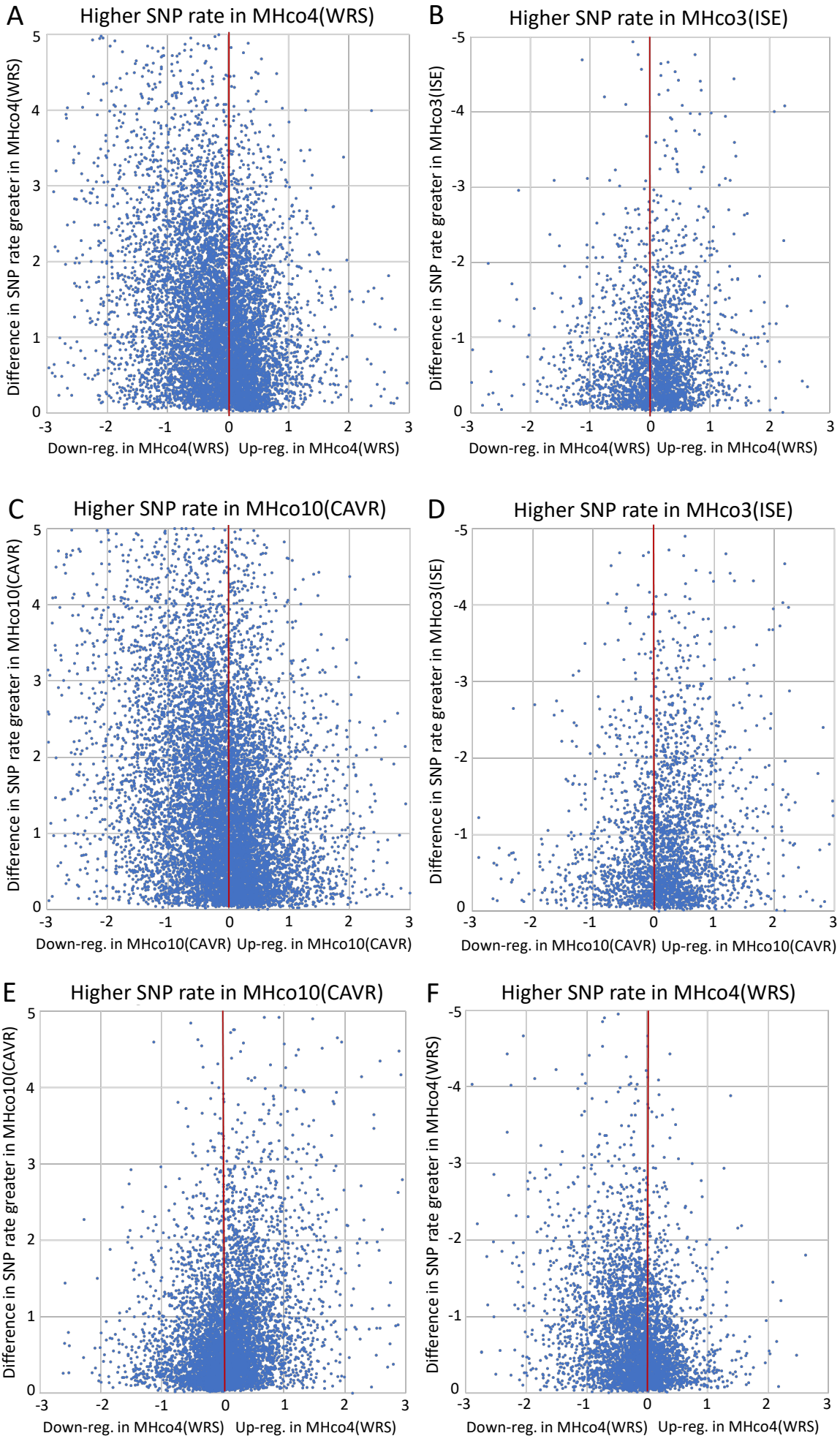

A
Higher SNP rate in MHco4(WRS)
Difference in SNP rate greater in MHco4(WRS)
Down-reg. in MHco4(WRS)
Up-reg. in MHco4(WRS)
B
Higher SNP rate in MHco3(ISE)
Difference in SNP rate greater in MHco3(ISE)
Down-reg. in MHco4(WRS)
Up-reg. in MHco4(WRS)
C
Higher SNP rate in MHco10(CAVR)
Difference in SNP rate greater in MHco10(CAVR)
Down-reg. in MHco10(CAVR)
Up-reg. in MHco10(CAVR)
D
Higher SNP rate in MHco3(ISE)
Difference in SNP rate greater in MHco3(ISE)
Down-reg. in MHco10(CAVR)
Up-reg. in MHco10(CAVR)
E
Higher SNP rate in MHco10(CAVR)
Difference in SNP rate greater in MHco10(CAVR)
Down-reg. in MHco4(WRS)
Up-reg. in MHco4(WRS)
F
Higher SNP rate in MHco4(WRS)
Difference in SNP rate greater in MHco4(WRS)
Down-reg. in MHco4(WRS)
Up-reg. in MHco4(WRS)

Figure 4 of 5 pdf, (ppt available)

Figure 4

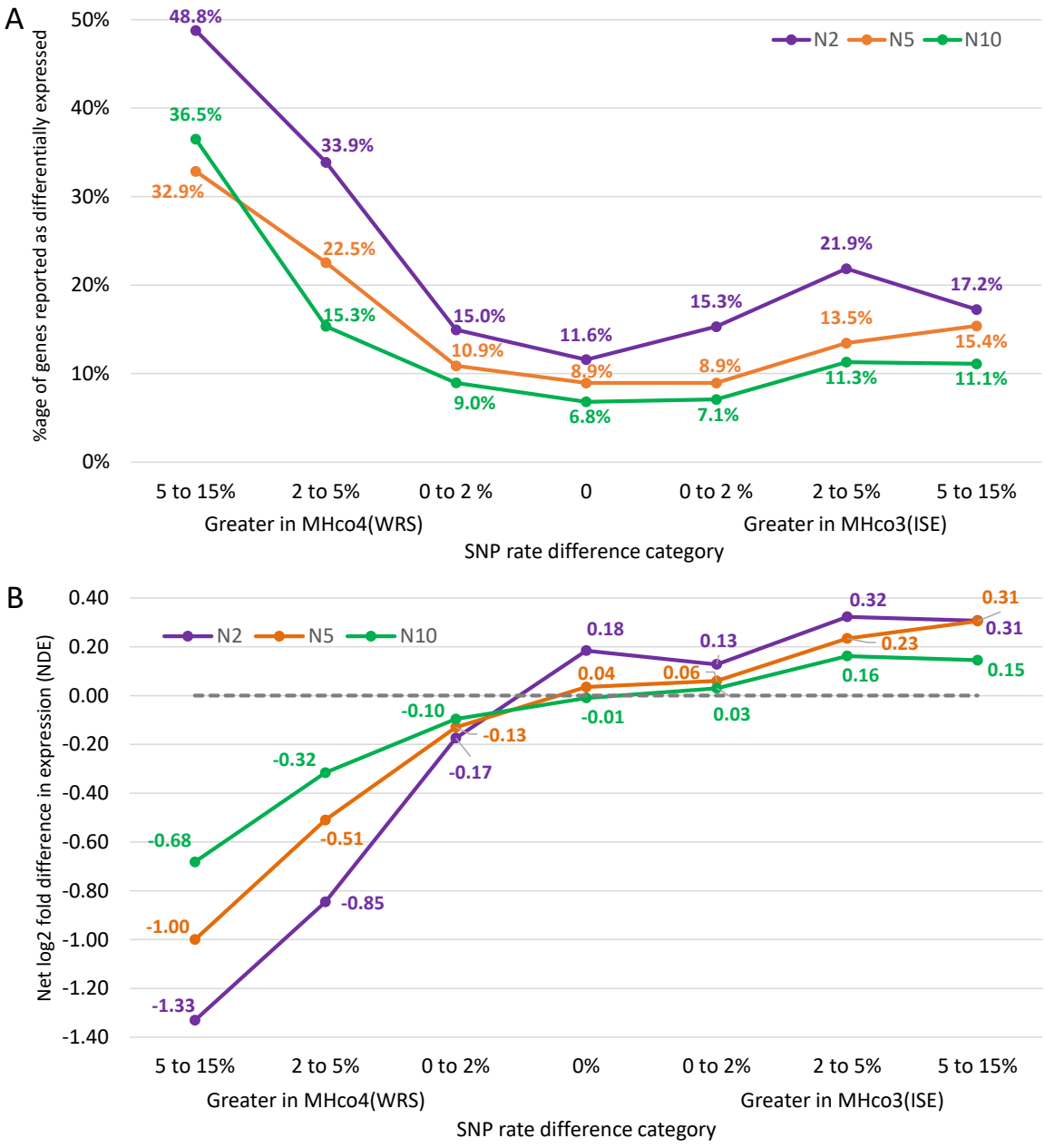
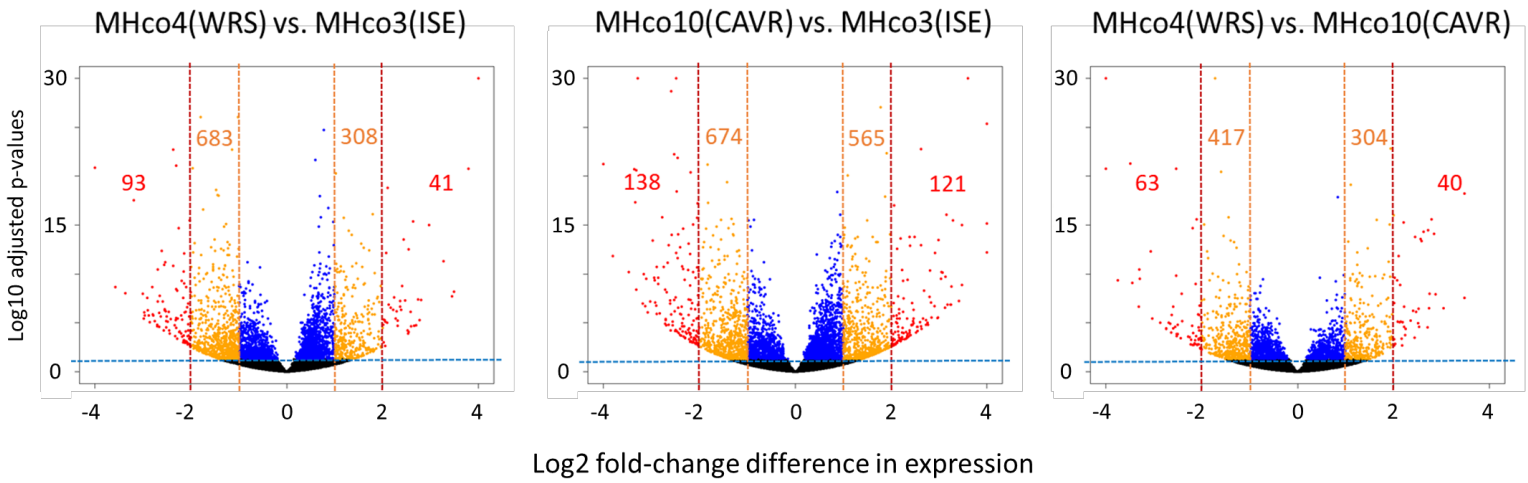
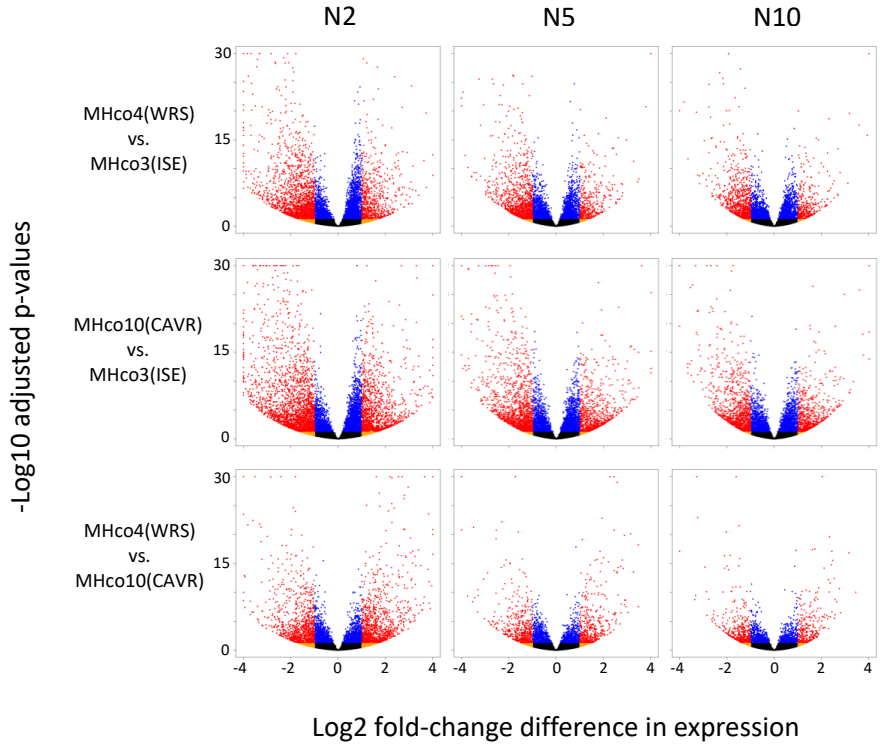


Figure 5 of 5 pdf, (ppt available)

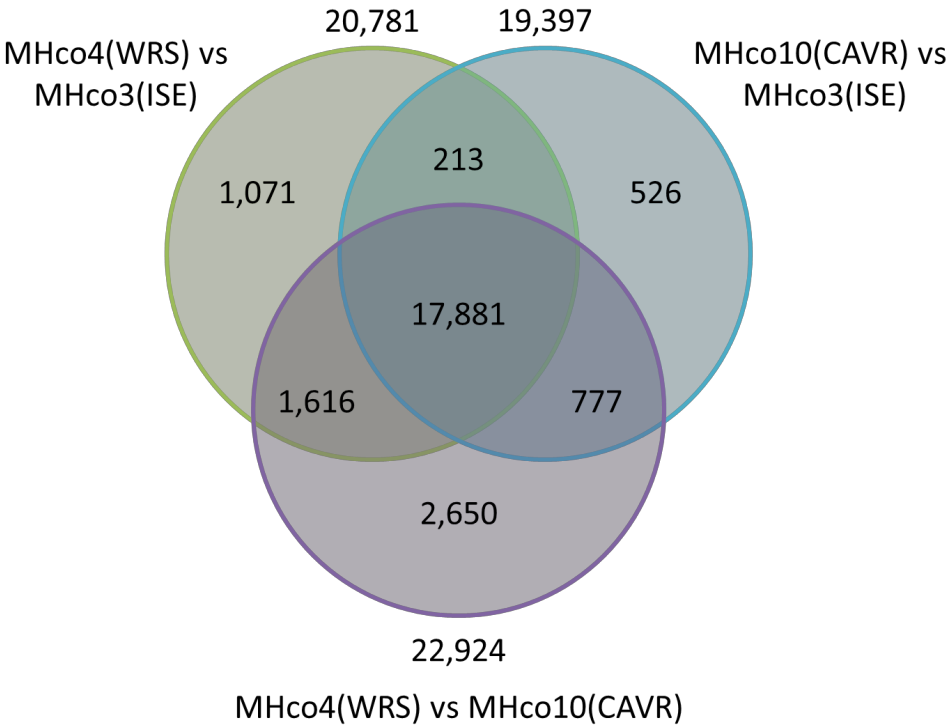
Figure 5



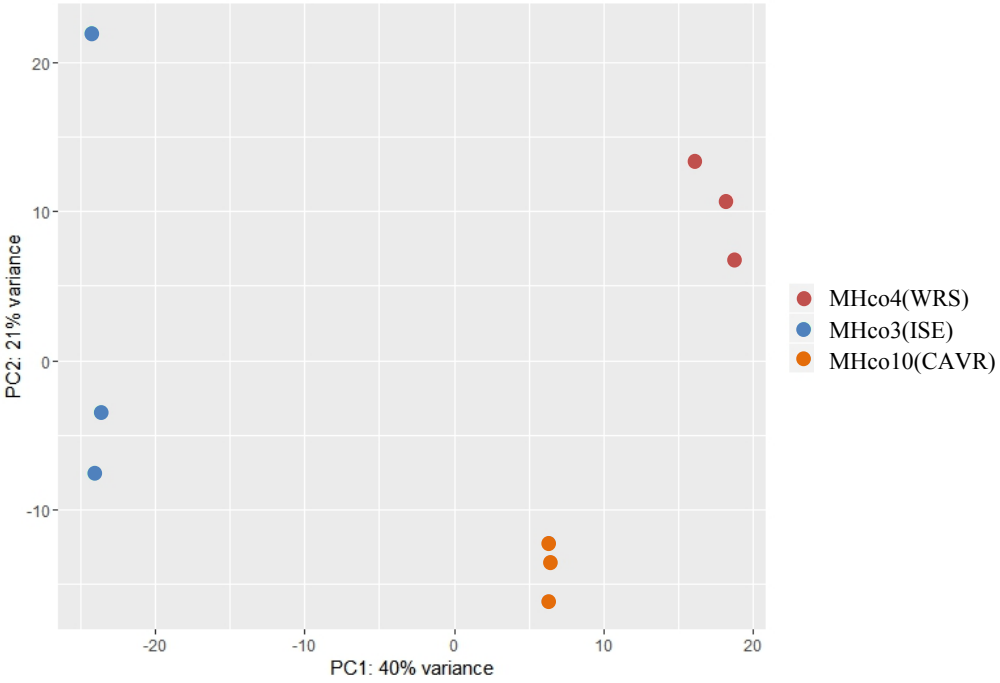
Supplementary Figure S1



Supplementary Figure S2



Supplementary Figure S3



Supplementary Figure S4